

Genomes of stigonematalean cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids

Tal Dagan^{1*}, Mayo Roettger^{2*}, Karina Stucken¹, Giddy Landan^{1,2}, Robin Koch¹, Peter Major², Sven B. Gould², Vadim V. Goremykin³, Rosmarie Rippka⁴, Nicole Tandeau de Marsac^{4,5}, Muriel Gugger⁶, Peter J. Lockhart⁷, John F. Allen^{8,9}, Iris Brune¹⁰, Irena Maus¹⁰, Alfred Pühler¹⁰, William F. Martin²

¹ Institute of Genomic Microbiology, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany

² Institute of Molecular Evolution, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany

³ IASMA Research and Innovation Center, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, (TN), Italy

⁴ Unité des Cyanobactéries, Institut Pasteur, CNRS URA 2172, 75724 Paris cedex 15, France

⁵ Present address: Aix-Marseille University, LCB, CNRS UPR 9043, 13402 Marseille, France

⁶ Laboratoire Collection des Cyanobactéries, Institut Pasteur, 75724 Paris cedex 15, France

⁷ Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

⁸ School of Biological and Chemical Sciences, Queen Mary, University of London, UK

⁹ Research Department of Genetics Evolution and Environment, University College London, UK

¹⁰ Center for Biotechnology, University of Bielefeld, Bielefeld, Germany

* These authors contributed equally to this work.

Corresponding author: Tal Dagan, email: tal.dagan@hhu.de.

Supplementary References.....	2
Supplementary Figure S1.....	2
Supplementary Figure S2.....	3
Supplementary Figure S3.....	4
Supplementary Figure S4.....	5
Supplementary Figure S5.....	6
Supplementary Figure S6.....	7
Supplementary Table S1.....	9
Supplementary Table S2.....	29
Supplementary Table S3.....	95

Supplementary References

Kaneshisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109-D114.

Tatusov RL, et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29:22-28.

The MathWorks Inc. 2012. MATLAB. Version: 7.14. (R2012a).

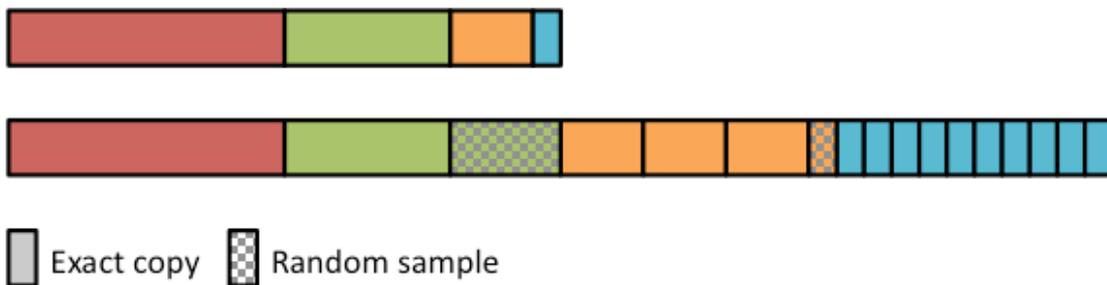


Fig. S1. Producing a weighted concatenated sequence alignment. In the common alignment concatenation procedure (top bar), a few long genes may contribute the majority of alignment columns to the resulting alignment. In our data, for example, 39% of the genes account for 61% of the columns. In weighted concatenation (bottom bar), each gene is over-sampled to the size of the longest gene, by concatenating exact repeats of its alignment, and the remainder being a random sample of columns (chequered blocks). This ensures that all genes contribute the same number of columns to the concatenated alignment.

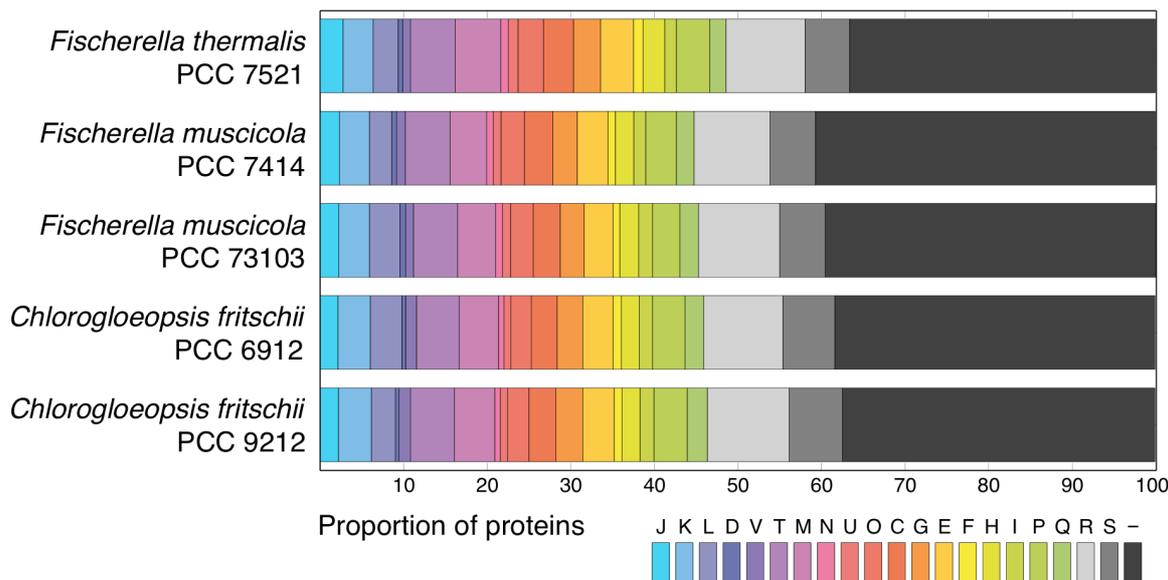


Fig. S2. Functional classification of proteins in genomes of Subsection V cyanobacteria.

CDSs in each sequenced cyanobacterium were classified into functional categories according to the COG functional classification (Tatusov et al. 2001). Functional categories are colored according to the one letter code: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; D, cell cycle control, cell division, chromosomal partitioning; V, defence mechanisms; T, signal transduction; M, cell wall/membrane/envelope biosynthesis; N, cell motility; Z, cytoskeleton; U, intracellular trafficking, secretion, vesicular transport; O, posttranslational modification, protein turnover, chaperones; C, energy production and conversion; transport and metabolism of: G, carbohydrate; E, amino acids; F, nucleotides; H, coenzymes; I, lipids; P, inorganic ions; Q, secondary metabolites. R, general function prediction only; S, function unknown; -, no COG annotation.

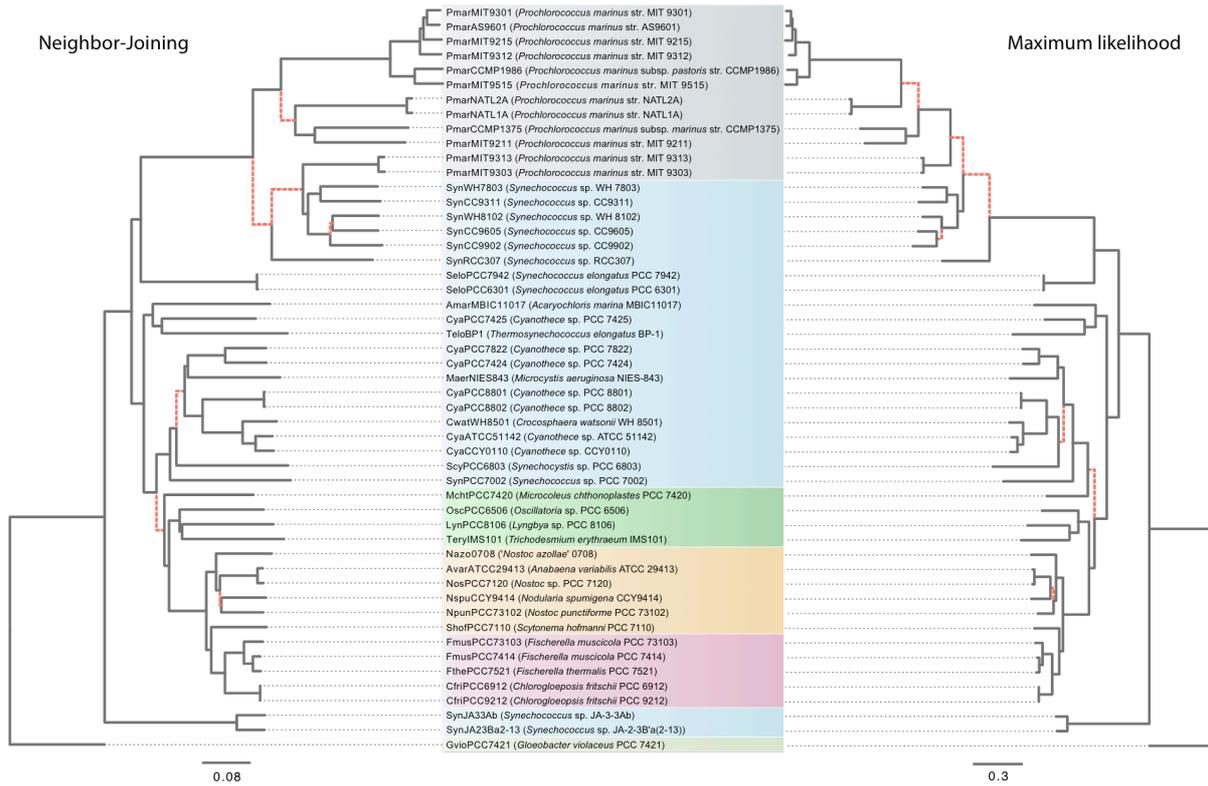


Fig. S3. Maximum likelihood and neighbor-joining phylogenies reconstructed from the concatenated alignment of 324 universal cyanobacterial single-copy proteins. The trees are rooted by *Gloeobacter violaceus* PCC 7421. Bootstrap support on neighbor-joining tree was 100 % for all branches. Incompatible branches between neighbor-joining and maximum likelihood phylogeny highlighted as dashed red lines. Prochlorales (grey), Chroococcales (blue), Oscillatoriales (green), Nostocales (orange), Stigonematales (red), Gloeobacterales (light green).

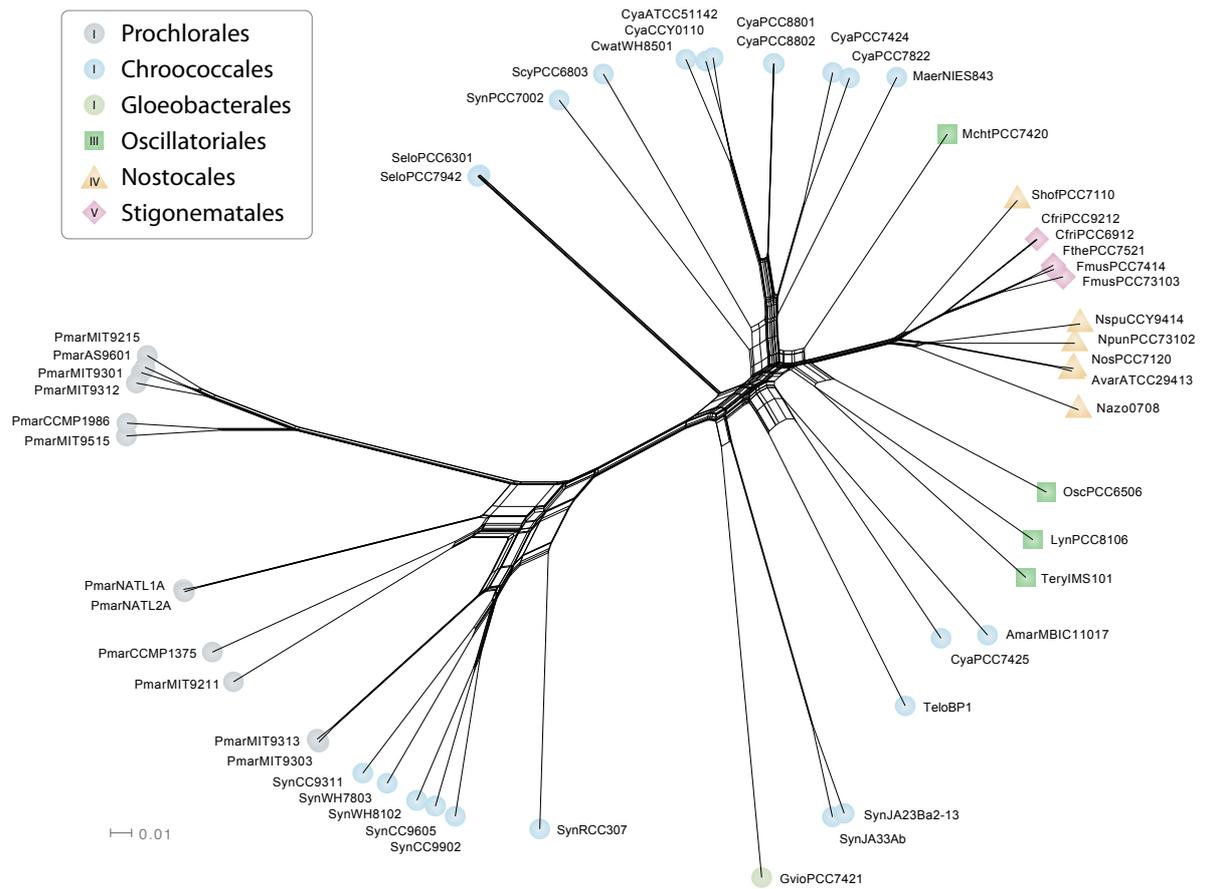


Fig. S4. NeighborNet reconstructed from weighted concatenated alignment of 324 universal cyanobacterial single-copy proteins.

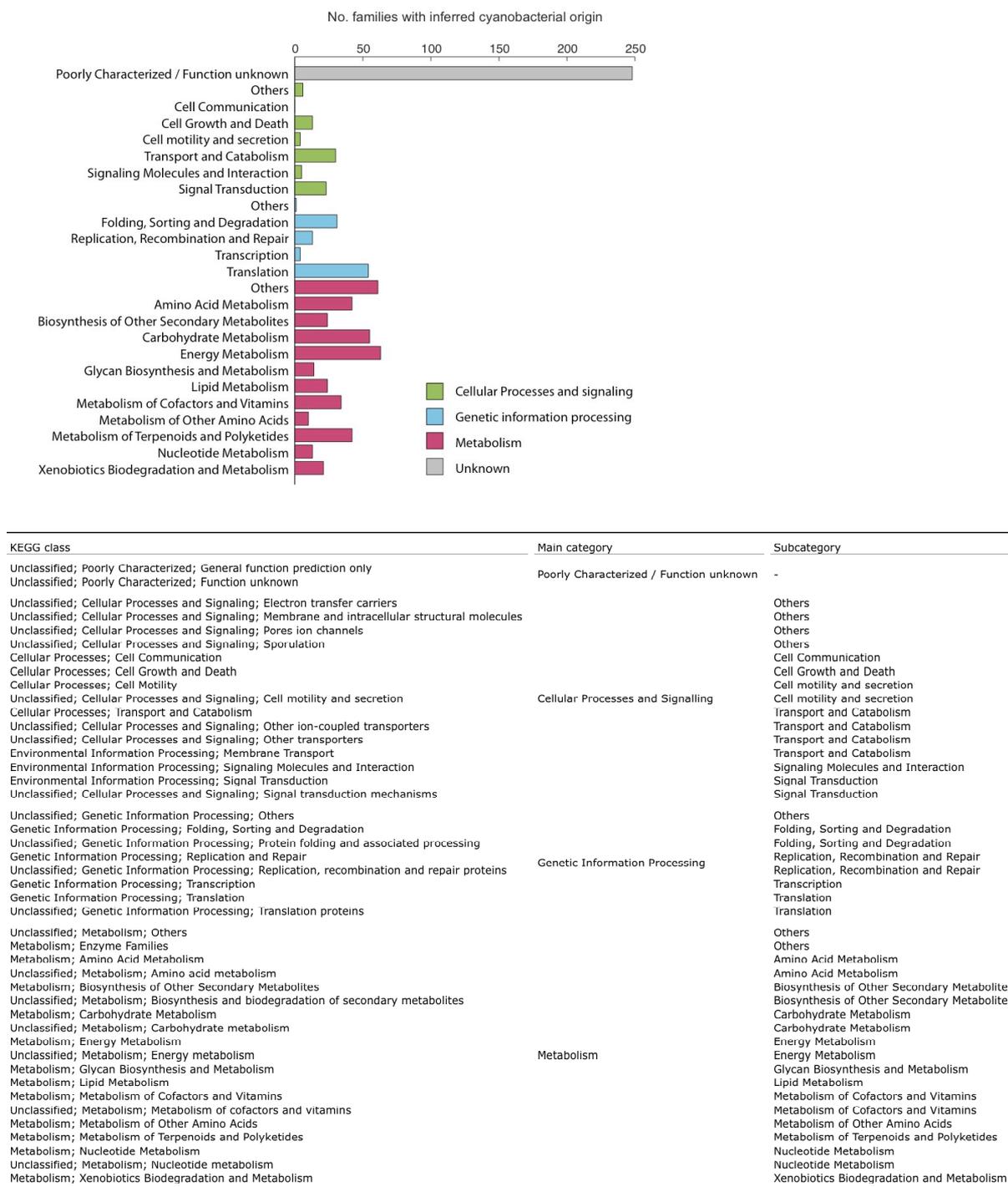


Fig. S5. Functional categories of plant protein families of cyanobacterial origin. Functional categories of CDSs were inferred by protein sequence similarity search in KEGG database (Kaneshisa *et al.* 2012). Functional main categories were assigned according to the annotation of the best blast hit in the database with E-value $\leq 10^{-10}$ and identity of at least 25%. Protein families were classified into the functional categories according to the most abundant functional assignment of their member proteins.

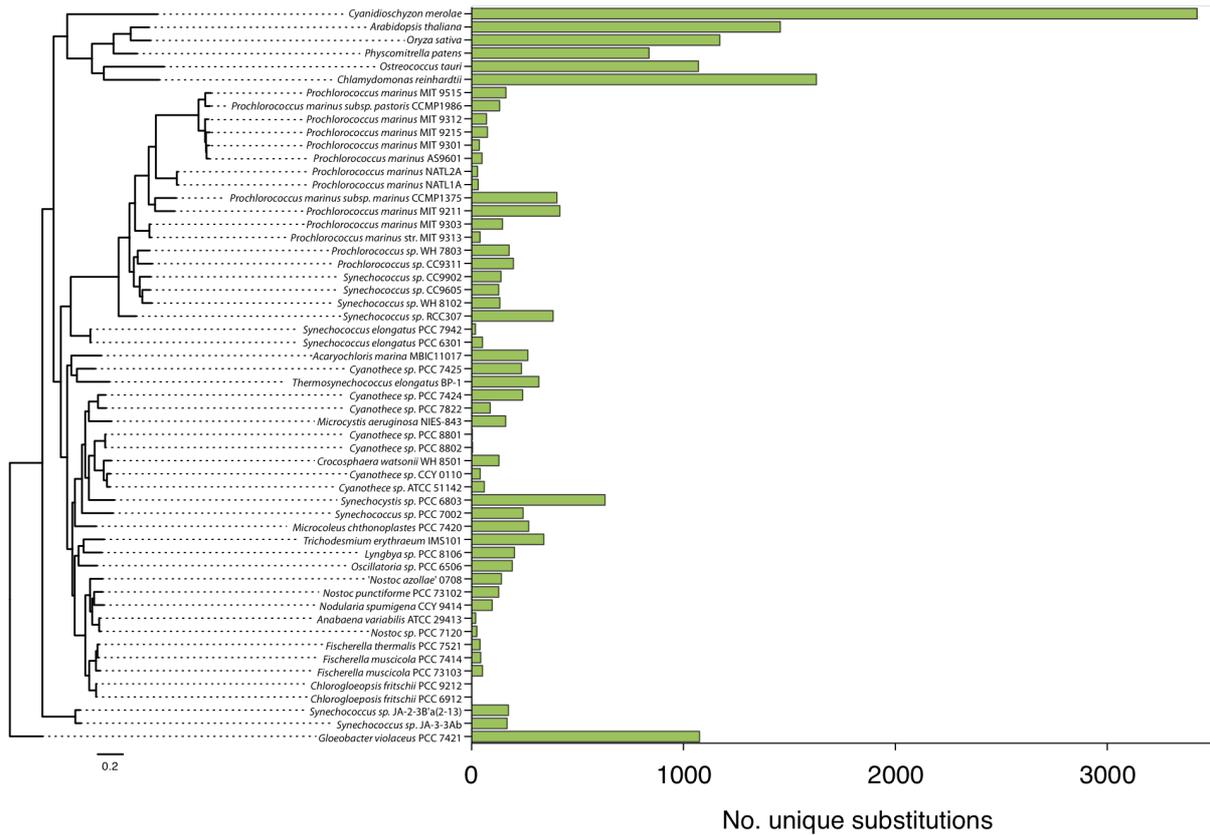


Fig. S6. Topology and distribution of unique characters in a maximum likelihood phylogenetic tree reconstructed from universal cyanobacterial genes and plant nuclear genes of cyanobacterial origin. The frequency of unique substitutions per organism was calculated from the concatenated alignment of 23 universal protein families common to cyanobacteria and plants. Amino acid position is counted as unique if the character state at that position is unique to a particular organism. The tree is rooted by *Gloeobacter violaceus* PCC 7421.

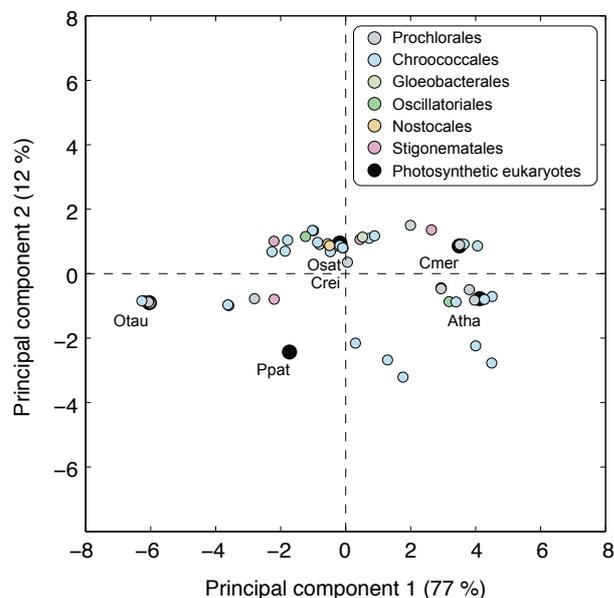


Fig. S7. Principal component analysis of amino acid usage within 611 proteins of endosymbiotic origin. Average amino acid frequencies for respective amino acid sequences for each cyanobacterium or photosynthetic eukaryote were used to perform a principal component analysis using princomp function of MATLAB (The Mathworks Inc. 2012). The first two principal components are shown. Sequences of photosynthetic eukaryotes are indicated as black circles. Atha: *Arabidopsis thaliana*, Cmer: *Cyanidioschyzon merolae*, Crei: *Chlamydomonas reinhardtii*, Osat: *Oryza sativa*, Otau: *Ostreococcus tauri*, Ppat: *Physcomitrella patens*.

Supp Table S1.

Protein families identified as unique and common to all filamentous cyanobacteria (Subsections III, IV and V). Best BLAST hits in non-filamentous cyanobacteria, and outside cyanobacteria are also cited.

Protein identifiers	PCC 6912	UYC_02149
	PCC 9212	Chlorogloopsis_PCC9212_joined_1436
	PCC 73103	FisPCC73103_5191
	PCC 7414	FisPCC7414_5731
	PCC 7521	Fischerella_sp._PCC7521_562
Protein length		291
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	Delta proteobacteria
	Organism	delta proteobacterium NaphS2
	Accession	ref ZP_07199851.1
	E-value	1.00E-107
	BLAST percent identity	53
	BLAST alignment length	281
Best hit in non-filamentous cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. WH 5701
	Accession	ref ZP_01085346.1
	E-value	9.00E-119
	BLAST percent identity	63
	BLAST alignment length	258

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_00415
	PCC 9212	Chlorogloopsis_PCC9212_joined_6496
	PCC 73103	FisPCC73103_1902 FisPCC73103_5577
	PCC 7414	FisPCC7414_2377
	PCC 7521	Fischerella_sp._PCC7521_3136
Protein length		122
Putative gene product		Anti-sigma F factor antagonist AltName: Full=Stage II sporulation protein AA;
Protein name		-
COG-number		COG1366
Best hit outside cyanobacteria	Phylum	-
	Organism	Gordonia terrae NBRC 100016
	Accession	ref ZP_09797653.1
	E-value	4.00E-14
	BLAST percent identity	42
	BLAST alignment length	93
Best hit in non-filamentous cyanobacteria	Organism	Microcystis aeruginosa NIES-843
	Accession	ref YP_001660919.1
	E-value	2.00E-11
	BLAST percent identity	36
	BLAST alignment length	89
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05037371.1
	E-value	6.00E-42
	BLAST percent identity	69
	BLAST alignment length	103

Protein identifiers	PCC 6912	UYC_07524
	PCC 9212	Chlorogloopsis_PCC9212_joined_3976

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 73103	FisPCC73103_1525
	PCC 7414	FisPCC7414_1514
	PCC 7521	Fischerella_sp._PCC7521_923
Protein length		398
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002376549.1
	E-value	3.00E-10
	BLAST percent identity	32
	BLAST alignment length	239
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05035642.1
	E-value	2.00E-119
	BLAST percent identity	50
	BLAST alignment length	369

Protein identifiers	PCC 6912	UYC_05040
	PCC 9212	Chlorogloopsis_PCC9212_joined_675
	PCC 73103	FisPCC73103_3432
	PCC 7414	FisPCC7414_5904
	PCC 7521	Fischerella_sp._PCC7521_3400

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		273
Putative gene product		Platelet-activating factor acetylhydrolase IB subunit beta EC=3.1.1.47; AltName: Full=PAF acetylhydrolase 30 kDa subunit; Short=PAF-AH 30 kDa subunit; AltName: Full=PAF-AH subunit beta; AltName: Full=PAFAH subunit beta; AltName: Full=Platelet-activating factor acetylhydrolase alpha 2 subunit; Short=PAF-AH alpha 2;
Protein name		EstA
COG-number		COG2755
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Clostridium kluyveri DSM 555
	Accession	ref YP_001395014.1
	E-value	5.00E-30
	BLAST percent identity	38
	BLAST alignment length	180
Best hit in non-filamentous cyanobacteria	Organism	Synechococcus sp. PCC 7002
	Accession	ref YP_001733839.1
	E-value	6.00E-54
	BLAST percent identity	50
	BLAST alignment length	200
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02255
	PCC 9212	Chlorogloopsis_PCC9212_joined_1533
	PCC 73103	FisPCC73103_3697
	PCC 7414	FisPCC7414_4564
	PCC 7521	Fischerella_sp._PCC7521_4367
Protein length		361

Dagan, Roettger, *et al.* – Supplementary Material

Putative gene product		Spermidine/putrescine-binding periplasmic protein 1 Short=SPBP; Flags: Precursor;
Protein name		PotD
COG-number		COG0687
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Peptostreptococcus anaerobius 653-L
	Accession	ref ZP_06424692.1
	E-value	1.00E-79
	BLAST percent identity	37
	BLAST alignment length	346
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002376537.1
	E-value	1.00E-68
	BLAST percent identity	37
	BLAST alignment length	319
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01657
	PCC 9212	Chlorogloopsis_PCC9212_joined_982
	PCC 73103	FisPCC73103_5115
	PCC 7414	FisPCC7414_2556
	PCC 7521	Fischerella_sp._PCC7521_5275
Protein length		281
Putative gene product		Band 7 protein CG32245
Protein name		-
COG-number		COG0330

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	Gamma proteobacteria
	Organism	Pseudoalteromonas tunicata D2
	Accession	ref ZP_01135473.1
	E-value	4.00E-55
	BLAST percent identity	39
	BLAST alignment length	298
Best hit in non-filamentous cyanobacteria	Organism	Synechococcus sp. JA-2-3B'a(2-13)
	Accession	ref YP_476585.1
	E-value	7.00E-82
	BLAST percent identity	44
	BLAST alignment length	292
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05036852.1
	E-value	3.00E-73
	BLAST percent identity	45
	BLAST alignment length	249

Protein identifiers	PCC 6912	UYC_00563
	PCC 9212	Chlorogloopsis_PCC9212_joined_6635
	PCC 73103	FisPCC73103_3681
	PCC 7414	FisPCC7414_5241
	PCC 7521	Fischerella_sp._PCC7521_3251
Protein length		156
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7425
	Accession	ref YP_002485010.1
	E-value	3.00E-18
	BLAST percent identity	40
	BLAST alignment length	121
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCMEE 5410
	Accession	ref ZP_09248396.1
	E-value	2.00E-12
	BLAST percent identity	33
	BLAST alignment length	119

Protein identifiers	PCC 6912	UYC_05545
	PCC 9212	Chlorogloopsis_PCC9212_joined_6970
	PCC 73103	FisPCC73103_4012
	PCC 7414	FisPCC7414_6259
	PCC 7521	Fischerella_sp._PCC7521_3873
Protein length		514
Putative gene product		putative beta-lactamase
Protein name		AmpC
COG-number		COG2367
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Brevibacillus brevis NBRC 100599
	Accession	ref YP_002774101.1
	E-value	5.00E-29
	BLAST percent identity	29
	BLAST alignment length	296
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	ref YP_002375824.1
	E-value	1.00E-109
	BLAST percent identity	58
	BLAST alignment length	281
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08975945.1
	E-value	1.00E-93
	BLAST percent identity	52
	BLAST alignment length	275

Protein identifiers	PCC 6912	UYC_01446
	PCC 9212	Chlorogloopsis_PCC9212_joined_5058
	PCC 73103	FisPCC73103_1595
	PCC 7414	FisPCC7414_5951
	PCC 7521	Fischerella_sp._PCC7521_1609
Protein length		218
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Dagan, Roettger, *et al.* – Supplementary Material

Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05037861.1
	E-value	6.00E-42
	BLAST percent identity	41
	BLAST alignment length	218

Protein identifiers	PCC 6912	UYC_03536
	PCC 9212	Chlorogloopsis_PCC9212_joined_5618
	PCC 73103	FisPCC73103_1802
	PCC 7414	FisPCC7414_1385
	PCC 7521	Fischerella_sp._PCC7521_3030
Protein length		583
Putative gene product		serine/threonine kinase
Protein name		-
COG-number		COG0515
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Crocospaera watsonii WH 8501
	Accession	ref ZP_00516977.1
	E-value	7.00E-66
	BLAST percent identity	43
	BLAST alignment length	310
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCMEE 5410
	Accession	ref ZP_09249476.1
	E-value	4.00E-62

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	40
	BLAST alignment length	336

Protein identifiers	PCC 6912	UYC_03071
	PCC 9212	Chlorogloopsis_PCC9212_joined_5934
	PCC 73103	FisPCC73103_7654
	PCC 7414	FisPCC7414_1793
	PCC 7521	Fischerella_sp._PCC7521_1971
Protein length		214
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002377363.1
	E-value	2.00E-22
	BLAST percent identity	33
	BLAST alignment length	211
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05038021.1
	E-value	2.00E-22
	BLAST percent identity	33
	BLAST alignment length	198

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_06488
	PCC 9212	Chlorogloopsis_PCC9212_joined_3020
	PCC 73103	FisPCC73103_3333
	PCC 7414	FisPCC7414_1107
	PCC 7521	Fischerella_sp._PCC7521_4792
Protein length		95
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 8801
	Accession	ref YP_002374341.1
	E-value	1.00E-20
	BLAST percent identity	47
	BLAST alignment length	95
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08971275.1
	E-value	4.00E-16
	BLAST percent identity	43
	BLAST alignment length	92

Protein identifiers	PCC 6912	UYC_04051
	PCC 9212	Chlorogloopsis_PCC9212_joined_7263
	PCC 73103	FisPCC73103_1401
	PCC 7414	FisPCC7414_2624

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_4221
Protein length		281
Putative gene product		Chaperone protein clpB
Protein name		-
COG-number		COG0013
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 8801
	Accession	ref YP_002371667.1
	E-value	4.00E-16
	BLAST percent identity	37
	BLAST alignment length	253
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08971398.1
	E-value	8.00E-11
	BLAST percent identity	34
	BLAST alignment length	251

Protein identifiers	PCC 6912	UYC_01445
	PCC 9212	Chlorogloopsis_PCC9212_joined_5059
	PCC 73103	FisPCC73103_1594
	PCC 7414	FisPCC7414_5952
	PCC 7521	Fischerella_sp._PCC7521_1610
Protein length		292
Putative gene product		hypothetical protein

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		COG2199
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05388
	PCC 9212	Chlorogloopsis_PCC9212_joined_7695
	PCC 73103	FisPCC73103_3927
	PCC 7414	FisPCC7414_698
	PCC 7521	Fischerella_sp._PCC7521_1750
Protein length		397
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-

	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7822
	Accession	ref YP_003888968.1
	E-value	3.00E-12
	BLAST percent identity	35
	BLAST alignment length	185
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05035642.1
	E-value	2.00E-138
	BLAST percent identity	56
	BLAST alignment length	369

Protein identifiers	PCC 6912	UYC_04188
	PCC 9212	Chlorogloopsis_PCC9212_joined_4760
	PCC 73103	FisPCC73103_3633
	PCC 7414	FisPCC7414_6840
	PCC 7521	Fischerella_sp._PCC7521_2784
Protein length		725
Putative gene product		Probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SPINDLY EC=2.4.1.-;
Protein name		-
COG-number		COG0457
Best hit outside cyanobacteria	Phylum	Thermotogae
	Organism	Petrotoga mobilis SJ95
	Accession	ref YP_001568946.1
	E-value	1.00E-26

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	32
	BLAST alignment length	303
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002379375.1
	E-value	2.00E-29
	BLAST percent identity	32
	BLAST alignment length	350
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05038442.1
	E-value	2.00E-54
	BLAST percent identity	27
	BLAST alignment length	760

Protein identifiers	PCC 6912	UYC_03604
	PCC 9212	Chlorogloopsis_PCC9212_joined_5555
	PCC 73103	FisPCC73103_6292
	PCC 7414	FisPCC7414_1052
	PCC 7521	Fischerella_sp._PCC7521_2087
Protein length		1015
Putative gene product		ADP,ATP carrier protein 1 AltName: Full=ADP/ATP translocase 1;
Protein name		-
COG-number		COG3202
Best hit outside cyanobacteria	Phylum	Bacterioidetes
	Organism	Croceibacter atlanticus HTCC2559
	Accession	ref YP_003716609.1
	E-value	4.00E-09
	BLAST percent identity	26
	BLAST alignment length	332
Best hit in non-filamentous cyanobacteria	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02128
	PCC 9212	Chlorogloopsis_PCC9212_joined_1415
	PCC 73103	FisPCC73103_3830
	PCC 7414	FisPCC7414_5507
	PCC 7521	Fischerella_sp._PCC7521_291
Protein length		853
Putative gene product		heterocyst differentiation protein
Protein name		HetF
COG-number		COG4995
Best hit outside cyanobacteria	Phylum	Actinobacteria
	Organism	Actinoplanes missouriensis 431
	Accession	ref YP_005463751.1
	E-value	1.00E-27
	BLAST percent identity	30
	BLAST alignment length	327
Best hit in non-filamentous cyanobacteria	Organism	Trichodesmium erythraeum IMS101
	Accession	ref YP_723693.1
	E-value	1.00E-127
	BLAST percent identity	54
	BLAST alignment length	399

Dagan, Roettger, *et al.* – Supplementary Material

Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05037316.1
	E-value	5.00E-128
	BLAST percent identity	53
	BLAST alignment length	396

Protein identifiers	PCC 6912	UYC_05419
	PCC 9212	Chlorogloopsis_PCC9212_joined_7724
	PCC 73103	FisPCC73103_5343
	PCC 7414	FisPCC7414_4695
	PCC 7521	Fischerella_sp._PCC7521_1641
Protein length		779
Putative gene product		unknown protein
Protein name		HtpX
COG-number		COG0501
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Clostridium tetani E88
	Accession	ref NP_781800.1
	E-value	3.00E-25
	BLAST percent identity	30
	BLAST alignment length	325
Best hit in non-filamentous cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002377167.1
	E-value	2.00E-86
	BLAST percent identity	32
	BLAST alignment length	652
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05035421.1
	E-value	3.00E-103

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	31
	BLAST alignment length	803

Protein identifiers	PCC 6912	UYC_06779
	PCC 9212	Chlorogloopsis_PCC9212_joined_3287
	PCC 73103	FisPCC73103_4524
	PCC 7414	FisPCC7414_5772
	PCC 7521	Fischerella_sp._PCC7521_4517
Protein length		150
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001516592.1
	E-value	1.00E-38
	BLAST percent identity	50
	BLAST alignment length	139
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCMEE 5410
	Accession	ref ZP_09252623.1
	E-value	1.00E-39
	BLAST percent identity	50
	BLAST alignment length	139

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_04049
	PCC 9212	Chlorogloopsis_PCC9212_joined_7265
	PCC 73103	FisPCC73103_1399
	PCC 7414	FisPCC7414_2621
	PCC 7521	Fischerella_sp._PCC7521_4219
Protein length		183
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG2165
Best hit outside cyanobacteria	Phylum	Delta-proteobacteria
	Organism	Syntrophus aciditrophicus SB
	Accession	ref YP_462385.1
	E-value	1.00E-09
	BLAST percent identity	40
	BLAST alignment length	75
Best hit in non-filamentous cyanobacteria	Organism	Synechococcus sp. JA-3-3Ab
	Accession	ref ZP_07109834.1
	E-value	2.00E-13
	BLAST percent identity	33
	BLAST alignment length	156
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05037970.1
	E-value	5.00E-47
	BLAST percent identity	44
	BLAST alignment length	179

Protein identifiers	PCC 6912	UYC_05389
	PCC 9212	Chlorogloopsis_PCC9212_joined_7696
	PCC 73103	FisPCC73103_3928
	PCC 7414	FisPCC7414_697

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_1749
Protein length		323
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-filamentous cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Hits in non-filamentous cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05035048.1
	E-value	1.00E-18
	BLAST percent identity	31
	BLAST alignment length	253

Supp Table S2.

Protein families identified as specific for heterocyst forming cyanobacteria (Subsections IV and V). Best BLAST hits in non-heterocyst forming cyanobacteria, and outside cyanobacteria are also cited.

Protein identifiers	PCC 6912	UYC_04554
	PCC 9212	Chlorogloopsis_PCC9212_joined_7170 Chlorogloopsis_PCC9212_joined_5881 Chlorogloopsis_PCC9212_joined_322
	PCC 73103	FisPCC73103_6274 FisPCC73103_4511 FisPCC73103_6338 FisPCC73103_3093
	PCC 7414	FisPCC7414_2761 FisPCC7414_1010 FisPCC7414_7362
	PCC 7521	Fischerella_sp._PCC7521_3704 Fischerella_sp._PCC7521_2900 Fischerella_sp._PCC7521_2199
Protein length		50
Putative gene product		helix-turn-helix protein, CopG
Protein name		copG
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

dataset	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_07797
	PCC 9212	Chlorogloopsis_PCC9212_joined_4230 Chlorogloopsis_PCC9212_joined_6628 Chlorogloopsis_PCC9212_joined_6625
	PCC 73103	FisPCC73103_4845
	PCC 7414	FisPCC7414_4842
	PCC 7521	Fischerella_sp._PCC7521_3090 Fischerella_sp._PCC7521_4841 Fischerella_sp._PCC7521_4840
Protein length		419
Putative gene product		hypothetical protein
Protein name		UGCG
COG-number		COG1215
Best hit outside cyanobacteria	Phylum	Planctomycetes
	Organism	Pirellula staleyi DSM 6068
	Accession	ref NP_868998.1
	E-value	4.00E-90
	BLAST percent identity	41
	BLAST alignment length	402
Best hit in non-heterocyst forming cyanobacteria	Organism	Synechococcus sp. CC9311
	Accession	ref YP_729392.1
	E-value	5.00E-29
	BLAST percent identity	36
	BLAST alignment length	251
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechocystis sp. PCC 6803 substr. GT-I
	Accession	ref YP_005383308.1
	E-value	7.00E-07

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	26
	BLAST alignment length	221

Protein identifiers	PCC 6912	UYC_04914
	PCC 9212	Chlorogloopsis_PCC9212_joined_555 Chlorogloopsis_PCC9212_joined_7637
	PCC 73103	FisPCC73103_1456
	PCC 7414	FisPCC7414_2145
	PCC 7521	Fischerella_sp._PCC7521_3306
Protein length		319
Putative gene product		glycosyl transferase, group 2 family protein
Protein name		-
COG-number		COG0463
Best hit outside cyanobacteria	Phylum	Bacteroidetes
	Organism	Rhodothermus marinus DSM 4252
	Accession	ref YP_003290430.1
	E-value	8.00E-45
	BLAST percent identity	35
	BLAST alignment length	308
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. ATCC 51142
	Accession	ref YP_001804598.1
	E-value	1.00E-26
	BLAST percent identity	30
	BLAST alignment length	320
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427151.1
	E-value	1.00E-119
	BLAST percent identity	57
	BLAST alignment length	316

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_07796
	PCC 9212	Chlorogloopsis_PCC9212_joined_4229
	PCC 73103	FisPCC73103_4844
	PCC 7414	FisPCC7414_4841
	PCC 7521	Fischerella_sp._PCC7521_4839
Protein length		421
Putative gene product		glycosyl transferase, group 1 family protein
Protein name		-
COG-number		COG0438
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Bacillus licheniformis DSM 13 = ATCC 14580
	Accession	ref YP_080330.1
	E-value	5.00E-40
	BLAST percent identity	30
	BLAST alignment length	398
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01620616.1
	E-value	5.00E-40
	BLAST percent identity	29
	BLAST alignment length	418
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08429743.1
	E-value	2.00E-35
	BLAST percent identity	28
	BLAST alignment length	407

Protein identifiers	PCC 6912	UYC_05513
	PCC 9212	Chlorogloopsis_PCC9212_joined_6939
	PCC 73103	FisPCC73103_4190
	PCC 7414	FisPCC7414_1002

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_2890
Protein length		470
Putative gene product		Uncharacterized protein slr1819
Protein name		-
COG-number		COG1357
Best hit outside cyanobacteria	Phylum	Actinobacteria
	Organism	Streptomyces viridochromogenes DSM 40736
	Accession	ref ZP_07308783.1
	E-value	3.00E-37
	BLAST percent identity	33
	BLAST alignment length	376
Best hit in non-heterocyst forming cyanobacteria	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001521229.1
	E-value	2.00E-48
	BLAST percent identity	39
	BLAST alignment length	351
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCME 5410
	Accession	ref ZP_09251784.1
	E-value	3.00E-46
	BLAST percent identity	41
	BLAST alignment length	255

Protein identifiers	PCC 6912	UYC_06951
	PCC 9212	Chlorogloopsis_PCC9212_joined_3443
	PCC 73103	FisPCC73103_1818
	PCC 7414	FisPCC7414_2222
	PCC 7521	Fischerella_sp._PCC7521_5078
Protein length		804
Putative gene product		Tripartite motif-containing protein 71 AltName: Full=Lin-41 homolog;

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		COG2319
Best hit outside cyanobacteria	Phylum	beta-proteobacteria
	Organism	Burkholderia thailandensis TXDOH
	Accession	ref ZP_02370139.1
	E-value	5.00E-156
	BLAST percent identity	39
	BLAST alignment length	707
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427138.1
	E-value	2.00E-151
	BLAST percent identity	39
	BLAST alignment length	727

Protein identifiers	PCC 6912	UYC_05846
	PCC 9212	Chlorogloopsis_PCC9212_joined_2421
	PCC 73103	FisPCC73103_2218
	PCC 7414	FisPCC7414_747
	PCC 7521	Fischerella_sp._PCC7521_4532 Fischerella_sp._PCC7521_4531
Protein length		543
Putative gene product		SWI/SNF chromatin-remodeling complex subunit snf22 EC=3.6.1.-; AltName: Full=SWI/SNF complex subunit snf22; AltName: Full=ATP-dependent helicase snf22;
Protein name		-
COG-number		COG5022

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05026763.1
	E-value	2.00E-46
	BLAST percent identity	47
	BLAST alignment length	201
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08432212.1
	E-value	4.00E-43
	BLAST percent identity	47
	BLAST alignment length	193

Protein identifiers	PCC 6912	UYC_07318
	PCC 9212	Chlorogloopsis_PCC9212_joined_3786
	PCC 73103	FisPCC73103_2418
	PCC 7414	FisPCC7414_870
	PCC 7521	Fischerella_sp._PCC7521_3431
Protein length		211
Putative gene product		N-acetylmuramoyl-L-alanine amidase cwIA EC=3.5.1.28; AltName: Full=Cell wall hydrolase; AltName: Full=Autolysin; Flags: Precursor;
Protein name		-
COG-number		COG3409
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Paenibacillus sp. HGF7
	Accession	ref ZP_08511092.1

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	1.00E-13
	BLAST percent identity	46
	BLAST alignment length	81
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01619407.1
	E-value	1.00E-11
	BLAST percent identity	45
	BLAST alignment length	73
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08431099.1
	E-value	2.00E-10
	BLAST percent identity	47
	BLAST alignment length	72

Protein identifiers	PCC 6912	UYC_06039
	PCC 9212	Chlorogloopsis_PCC9212_joined_2603
	PCC 73103	FisPCC73103_4362
	PCC 7414	FisPCC7414_2109
	PCC 7521	Fischerella_sp._PCC7521_2970
Protein length		599
Putative gene product		serine/threonine kinase
Protein name		PknB
COG-number		COG0515
Best hit outside cyanobacteria	Phylum	Acidobacteria
	Organism	Candidatus Chloracidobacterium thermophilum B
	Accession	ref YP_004863306.1
	E-value	4.00E-41
	BLAST percent identity	38
	BLAST alignment length	275

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001517647.1
	E-value	3.00E-93
	BLAST percent identity	41
	BLAST alignment length	476
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08432048.1
	E-value	2.00E-121
	BLAST percent identity	37
	BLAST alignment length	590

Protein identifiers	PCC 6912	UYC_02520
	PCC 9212	Chlorogloopsis_PCC9212_joined_1787
	PCC 73103	FisPCC73103_7745
	PCC 7414	FisPCC7414_5858
	PCC 7521	Fischerella_sp._PCC7521_4426
Protein length		147
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05804
	PCC 9212	Chlorogloopsis_PCC9212_joined_2382
	PCC 73103	FisPCC73103_4586
	PCC 7414	FisPCC7414_6442
	PCC 7521	Fischerella_sp._PCC7521_2714
Protein length		356
Putative gene product		UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit EC=2.4.1.-; AltName: Full=O-linked N-acetylglucosamine transferase 110 kDa subunit; AltName: Full=O-GlcNAc transferase subunit p110;
Protein name		OGT
COG-number		COG0457
Best hit outside cyanobacteria	Phylum	Chlamydiae
	Organism	Chthoniobacter flavus Ellin42
	Accession	ref ZP_03131794.1
	E-value	2.00E-39
	BLAST percent identity	38
	BLAST alignment length	249
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05025923.1
	E-value	8.00E-57
	BLAST percent identity	43
	BLAST alignment length	268

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427719.1
	E-value	2.00E-55
	BLAST percent identity	41
	BLAST alignment length	274

Protein identifiers	PCC 6912	UYC_00561
	PCC 9212	Chlorogloopsis_PCC9212_joined_6633
	PCC 73103	FisPCC73103_3676
	PCC 7414	FisPCC7414_5238
	PCC 7521	Fischerella_sp._PCC7521_3249
Protein length		310
Putative gene product		hydrolase, alpha/beta fold family
Protein name		TesD
COG-number		COG0596
Best hit outside cyanobacteria	Phylum	Planctomycetes
	Organism	Rhodopirellula baltica SH 1
	Accession	ref NP_870949.1
	E-value	2.00E-32
	BLAST percent identity	29
	BLAST alignment length	299
Best hit in non-heterocyst forming cyanobacteria	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001517562.1
	E-value	7.00E-22
	BLAST percent identity	28
	BLAST alignment length	282
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCME5 5410
	Accession	ref ZP_09251436.1
	E-value	6.00E-22
	BLAST percent identity	28

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	282
--	------------------------	-----

Protein identifiers	PCC 6912	UYC_01717
	PCC 9212	Chlorogloopsis_PCC9212_joined_1038
	PCC 73103	FisPCC73103_3880
	PCC 7414	FisPCC7414_5864
	PCC 7521	Fischerella_sp._PCC7521_5328
Protein length		394
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08431644.1
	E-value	1.00E-71
	BLAST percent identity	38
	BLAST alignment length	402

Protein identifiers	PCC 6912	UYC_00493
---------------------	----------	-----------

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 9212	Chlorogloopsis_PCC9212_joined_6571
	PCC 73103	FisPCC73103_3266
	PCC 7414	FisPCC7414_3800
	PCC 7521	Fischerella_sp._PCC7521_1475
Protein length		186
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01045
	PCC 9212	Chlorogloopsis_PCC9212_joined_4551
	PCC 73103	FisPCC73103_2015
	PCC 7414	FisPCC7414_6966
	PCC 7521	Fischerella_sp._PCC7521_3779

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		136
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG1132
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. CCY0110
	Accession	ref ZP_01728657.1
	E-value	2.00E-08
	BLAST percent identity	54
	BLAST alignment length	46
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08971538.1
	E-value	1.00E-07
	BLAST percent identity	53
	BLAST alignment length	45

Protein identifiers	PCC 6912	UYC_00932
	PCC 9212	Chlorogloopsis_PCC9212_joined_4661
	PCC 73103	FisPCC73103_3131
	PCC 7414	FisPCC7414_2762
	PCC 7521	Fischerella_sp._PCC7521_3703
Protein length		731
Putative gene product		toxin secretion ABC transporter ATP-binding protein
Protein name		HlyB

Dagan, Roettger, *et al.* – Supplementary Material

COG-number		COG2274
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Abiotrophia defectiva ATCC 49176
	Accession	ref ZP_04451120.1
	E-value	0
	BLAST percent identity	48
	BLAST alignment length	666
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. CCY0110
	Accession	ref ZP_01730354.1
	E-value	6.00E-180
	BLAST percent identity	45
	BLAST alignment length	724
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08973154.1
	E-value	4.00E-179
	BLAST percent identity	45
	BLAST alignment length	724

Protein identifiers	PCC 6912	UYC_05306
	PCC 9212	Chlorogloopsis_PCC9212_joined_7617
	PCC 73103	FisPCC73103_4344
	PCC 7414	FisPCC7414_2970
	PCC 7521	Fischerella_sp._PCC7521_947
Protein length		291
Putative gene product		Transforming growth factor-beta-induced protein ig-h3 Short=Beta ig-h3; Flags: Precursor;
Protein name		-
COG-number		COG2335
Best hit outside cyanobacteria	Phylum	Planctomycetes
	Organism	Isosphaera pallida ATCC 43644

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	ref YP_004178668.1
	E-value	5.00E-45
	BLAST percent identity	59
	BLAST alignment length	141
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. PCC 7822
	Accession	ref YP_003890676.1
	E-value	5.00E-47
	BLAST percent identity	62
	BLAST alignment length	138
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08492506.1
	E-value	1.00E-51
	BLAST percent identity	67
	BLAST alignment length	135

Protein identifiers	PCC 6912	UYC_07828
	PCC 9212	Chlorogloopsis_PCC9212_joined_4260
	PCC 73103	FisPCC73103_3814
	PCC 7414	FisPCC7414_5534
	PCC 7521	Fischerella_sp._PCC7521_5565
Protein length		98
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. ATCC 51142
	Accession	ref YP_001803803.1
	E-value	5.00E-08
	BLAST percent identity	62
	BLAST alignment length	42
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08973122.1
	E-value	5.00E-08
	BLAST percent identity	62
	BLAST alignment length	42

Protein identifiers	PCC 6912	UYC_01789
	PCC 9212	Chlorogloopsis_PCC9212_joined_1109
	PCC 73103	FisPCC73103_4695
	PCC 7414	FisPCC7414_5374
	PCC 7521	Fischerella_sp._PCC7521_4978
Protein length		206
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05026357.1
	E-value	5.00E-31

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	50
	BLAST alignment length	129
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427753.1
	E-value	2.00E-26
	BLAST percent identity	51
	BLAST alignment length	109

Protein identifiers	PCC 6912	UYC_05797
	PCC 9212	Chlorogloopsis_PCC9212_joined_2377
	PCC 73103	FisPCC73103_4593
	PCC 7414	FisPCC7414_6449
	PCC 7521	Fischerella_sp._PCC7521_2721
Protein length		153
Putative gene product		UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase EC=2.4.1.-; AltName: Full=O-GlcNAc; AltName: Full=OGT;
Protein name		-
COG-number		COG0457
Best hit outside cyanobacteria	Phylum	-
	Organism	Truepera radiovictrix DSM 17093
	Accession	ref YP_003706428.1
	E-value	6.00E-10
	BLAST percent identity	36
	BLAST alignment length	104
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05031148.1
	E-value	9.00E-26
	BLAST percent identity	47
	BLAST alignment length	112

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427719.1
	E-value	4.00E-09
	BLAST percent identity	35
	BLAST alignment length	97

Protein identifiers	PCC 6912	UYC_04475
	PCC 9212	Chlorogloopsis_PCC9212_joined_7097
	PCC 73103	FisPCC73103_5386
	PCC 7414	FisPCC7414_4659
	PCC 7521	Fischerella_sp._PCC7521_3809
Protein length		80
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07109439.1
	E-value	2.00E-12
	BLAST percent identity	65
	BLAST alignment length	46
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08490986.1
	E-value	8.00E-09
	BLAST percent identity	59

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	46
--	------------------------	----

Protein identifiers	PCC 6912	UYC_04197
	PCC 9212	Chlorogloopsis_PCC9212_joined_4768
	PCC 73103	FisPCC73103_3625
	PCC 7414	FisPCC7414_6850
	PCC 7521	Fischerella_sp._PCC7521_2793
Protein length		534
Putative gene product		Cortactin-binding protein 2 Short=CortBP2;
Protein name		-
COG-number		COG0542
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05029654.1
	E-value	2.00E-31
	BLAST percent identity	43
	BLAST alignment length	200
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08426717.1
	E-value	2.00E-25
	BLAST percent identity	39
	BLAST alignment length	200

Protein identifiers	PCC 6912	UYC_00912
---------------------	----------	-----------

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 9212	Chlorogloopsis_PCC9212_joined_4678
	PCC 73103	FisPCC73103_6012
	PCC 7414	FisPCC7414_4979
	PCC 7521	Fischerella_sp._PCC7521_4522
Protein length		106
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_04632
	PCC 9212	Chlorogloopsis_PCC9212_joined_6854
	PCC 73103	FisPCC73103_2276
	PCC 7414	FisPCC7414_1448
	PCC 7521	Fischerella_sp._PCC7521_3488

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		78
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07111889.1
	E-value	6.00E-09
	BLAST percent identity	47
	BLAST alignment length	53
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08976064.1
	E-value	4.00E-07
	BLAST percent identity	54
	BLAST alignment length	46

Protein identifiers	PCC 6912	UYC_06363
	PCC 9212	Chlorogloopsis_PCC9212_joined_2900
	PCC 73103	FisPCC73103_5614
	PCC 7414	FisPCC7414_5956
	PCC 7521	Fischerella_sp._PCC7521_1613
Protein length		434
Putative gene product		Polycystic kidney disease protein 1-like 3 AltName: Full=Polycystin-1L3; AltName: Full=PC1-like 3 protein; Flags: Precursor;

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		COG1674
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02311
	PCC 9212	Chlorogloopsis_PCC9212_joined_1590
	PCC 73103	FisPCC73103_3905
	PCC 7414	FisPCC7414_715
	PCC 7521	Fischerella_sp._PCC7521_1765
Protein length		148
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-

Dagan, Roettger, *et al.* – Supplementary Material

	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05027134.1
	E-value	2.00E-07
	BLAST percent identity	36
	BLAST alignment length	118
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495444.1
	E-value	7.00E-08
	BLAST percent identity	35
	BLAST alignment length	117

Protein identifiers	PCC 6912	UYC_02385
	PCC 9212	Chlorogloopsis_PCC9212_joined_1664
	PCC 73103	FisPCC73103_4087
	PCC 7414	FisPCC7414_7684
	PCC 7521	Fischerella_sp._PCC7521_2825
Protein length		406
Putative gene product		T-cell immunomodulatory protein Short=Protein TIP; AltName: Full=Integrin-alpha FG-GAP repeat-containing protein 1; Flags: Precursor;
Protein name		-
COG-number		COG3209
Best hit outside cyanobacteria	Phylum	Delta-proteobacteria
	Organism	Geobacter uraniireducens Rf4
	Accession	ref YP_001231793.1
	E-value	7.00E-38

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	40
	BLAST alignment length	244
Best hit in non-heterocyst forming cyanobacteria	Organism	Gloeobacter violaceus PCC 7421
	Accession	ref NP_924006.1
	E-value	2.00E-52
	BLAST percent identity	41
	BLAST alignment length	283
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08426188.1
	E-value	8.00E-32
	BLAST percent identity	37
	BLAST alignment length	251

Protein identifiers	PCC 6912	UYC_04418
	PCC 9212	Chlorogloopsis_PCC9212_joined_4969
	PCC 73103	FisPCC73103_3292
	PCC 7414	FisPCC7414_772
	PCC 7521	Fischerella_sp._PCC7521_4554
Protein length		289
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3468
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming	Organism	Oscillatoria sp. PCC 6506

Dagan, Roettger, *et al.* – Supplementary Material

cyanobacteria	Accession	ref ZP_07112366.1
	E-value	6.00E-17
	BLAST percent identity	58
	BLAST alignment length	73
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset		
	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08491334.1
	E-value	9.00E-20
	BLAST percent identity	64
	BLAST alignment length	74

Protein identifiers	PCC 6912	UYC_02306
	PCC 9212	Chlorogloopsis_PCC9212_joined_1586
	PCC 73103	FisPCC73103_5822
	PCC 7414	FisPCC7414_2396
	PCC 7521	Fischerella_sp._PCC7521_5081
Protein length		
		289
Putative gene product		
		hypothetical protein
Protein name		
		-
COG-number		
		-
Best hit outside cyanobacteria		
	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria		
	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001518478.1
	E-value	4.00E-21
	BLAST percent identity	30
	BLAST alignment length	257

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCME 5410
	Accession	ref ZP_09252250.1
	E-value	4.00E-21
	BLAST percent identity	30
	BLAST alignment length	257

Protein identifiers	PCC 6912	UYC_04404
	PCC 9212	Chlorogloopsis_PCC9212_joined_4956
	PCC 73103	FisPCC73103_6611
	PCC 7414	FisPCC7414_6988
	PCC 7521	Fischerella_sp._PCC7521_2173
Protein length		227
Putative gene product		UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit EC=2.4.1.-; AltName: Full=O-linked N-acetylglucosamine transferase 110 kDa subunit; AltName: Full=O-GlcNAc transferase subunit p110;
Protein name		OGT
COG-number		COG0457
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Acetonema longum DSM 6540
	Accession	ref ZP_08622906.1
	E-value	2.00E-16
	BLAST percent identity	37
	BLAST alignment length	156
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05024606.1
	E-value	5.00E-39
	BLAST percent identity	50
	BLAST alignment length	165
Best hit in non-heterocyst forming cyanobacteria that are not in our	Organism	Microcoleus vaginatus FGP-2

Dagan, Roettger, *et al.* – Supplementary Material

dataset	Accession	ref ZP_08494216.1
	E-value	4.00E-37
	BLAST percent identity	47
	BLAST alignment length	167

Protein identifiers	PCC 6912	UYC_00189
	PCC 9212	Chlorogloopsis_PCC9212_joined_4440
	PCC 73103	FisPCC73103_4132
	PCC 7414	FisPCC7414_595
	PCC 7521	Fischerella_sp._PCC7521_2325
Protein length		73
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01619687.1
	E-value	1.00E-14
	BLAST percent identity	56
	BLAST alignment length	55
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Arthrospira maxima CS-328
	Accession	ref ZP_03272907.1
	E-value	6.00E-14
	BLAST percent identity	44
	BLAST alignment length	63

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_00936
	PCC 9212	Chlorogloopsis_PCC9212_joined_4657
	PCC 73103	FisPCC73103_6193
	PCC 7414	FisPCC7414_973
	PCC 7521	Fischerella_sp._PCC7521_1233
Protein length		498
Putative gene product		Leukotoxin secretion protein D
Protein name		lktD
COG-number		COG0845
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Eubacteriaceae bacterium ACC19a
	Accession	ref ZP_09321904.1
	E-value	2.00E-55
	BLAST percent identity	29
	BLAST alignment length	514
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01621292.1
	E-value	2.00E-34
	BLAST percent identity	28
	BLAST alignment length	468
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495096.1
	E-value	1.00E-31
	BLAST percent identity	26
	BLAST alignment length	506

Protein identifiers	PCC 6912	UYC_01625
	PCC 9212	Chlorogloopsis_PCC9212_joined_952

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 73103	FisPCC73103_3882
	PCC 7414	FisPCC7414_5866
	PCC 7521	Fischerella_sp._PCC7521_5330
Protein length		855
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3106
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08431646.1
	E-value	3.00E-131
	BLAST percent identity	35
	BLAST alignment length	883

Protein identifiers	PCC 6912	UYC_06613
	PCC 9212	Chlorogloopsis_PCC9212_joined_3134
	PCC 73103	FisPCC73103_6131
	PCC 7414	FisPCC7414_2127
	PCC 7521	Fischerella_sp._PCC7521_3287

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		434
Putative gene product		Peripheral-type benzodiazepine receptor-associated protein 1 AltName: Full=PRAX-1; AltName: Full=Peripheral benzodiazepine receptor-interacting protein; Short=PBR-IP; AltName: Full=RIMS-binding protein 1; Short=RIM-BP1;
Protein name		-
COG-number		COG3264
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Clostridium cellulolyticum H10
	Accession	ref YP_002505304.1
	E-value	5.00E-10
	BLAST percent identity	49
	BLAST alignment length	69
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05028499.1
	E-value	2.00E-09
	BLAST percent identity	26
	BLAST alignment length	185
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_07831
	PCC 9212	Chlorogloopsis_PCC9212_joined_4263
	PCC 73103	FisPCC73103_3817
	PCC 7414	FisPCC7414_5530
	PCC 7521	Fischerella_sp._PCC7521_5569
Protein length		413
Putative gene product		hypothetical protein

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		COG2217
Best hit outside cyanobacteria	Phylum	Delta-proteobacteria
	Organism	Sorangium cellulosum 'So ce 56'
	Accession	ref YP_001611891.1
	E-value	3.00E-05
	BLAST percent identity	40
	BLAST alignment length	65
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05031079.1
	E-value	1.00E-37
	BLAST percent identity	32
	BLAST alignment length	279
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05036455.1
	E-value	9.00E-09
	BLAST percent identity	52
	BLAST alignment length	54

Protein identifiers	PCC 6912	UYC_04851
	PCC 9212	Chlorogloopsis_PCC9212_joined_494
	PCC 73103	FisPCC73103_7554
	PCC 7414	FisPCC7414_3770
	PCC 7521	Fischerella_sp._PCC7521_1496
Protein length		65
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-

Dagan, Roettger, *et al.* – Supplementary Material

	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07110836.1
	E-value	8.00E-13
	BLAST percent identity	56
	BLAST alignment length	63
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08430235.1
	E-value	2.00E-12
	BLAST percent identity	58
	BLAST alignment length	55

Protein identifiers	PCC 6912	UYC_00685
	PCC 9212	Chlorogloopsis_PCC9212_joined_6753
	PCC 73103	FisPCC73103_5870
	PCC 7414	FisPCC7414_4272
	PCC 7521	Fischerella_sp._PCC7521_2514
Protein length		442
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05025165.1
	E-value	2.00E-60
	BLAST percent identity	49
	BLAST alignment length	225
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427412.1
	E-value	4.00E-61
	BLAST percent identity	46
	BLAST alignment length	241

Protein identifiers	PCC 6912	UYC_06095
	PCC 9212	Chlorogloopsis_PCC9212_joined_2655
	PCC 73103	FisPCC73103_1871
	PCC 7414	FisPCC7414_6465
	PCC 7521	Fischerella_sp._PCC7521_780
Protein length		410
Putative gene product		putative glycosyl transferase
Protein name		-
COG-number		COG0438
Best hit outside cyanobacteria	Phylum	Delta-proteobacteria
	Organism	delta proteobacterium NaphS2
	Accession	ref ZP_07199056.1
	E-value	2.00E-43
	BLAST percent identity	31
	BLAST alignment length	392
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05026852.1

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	6.00E-16
	BLAST percent identity	30
	BLAST alignment length	227
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechocystis sp. PCC 6803 substr. GT-I
	Accession	ref YP_005382114.1
	E-value	3.00E-15
	BLAST percent identity	31
	BLAST alignment length	219

Protein identifiers	PCC 6912	UYC_03938
	PCC 9212	Chlorogloopsis_PCC9212_joined_7370
	PCC 73103	FisPCC73103_2733
	PCC 7414	FisPCC7414_2813
	PCC 7521	Fischerella_sp._PCC7521_3173
Protein length		101
Putative gene product		Circadian clock protein kaiA
Protein name		KaiA
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07111746.1
	E-value	3.00E-39
	BLAST percent identity	64
	BLAST alignment length	99

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08491806.1
	E-value	2.00E-39
	BLAST percent identity	64
	BLAST alignment length	97

Protein identifiers	PCC 6912	UYC_06542
	PCC 9212	Chlorogloopsis_PCC9212_joined_3071
	PCC 73103	FisPCC73103_1786
	PCC 7414	FisPCC7414_7395
	PCC 7521	Fischerella_sp._PCC7521_4098
Protein length		240
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. PCC 7822
	Accession	ref YP_003885559.1
	E-value	5.00E-27
	BLAST percent identity	36
	BLAST alignment length	243
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechocystis sp. PCC 6803 substr. GT-I
	Accession	ref YP_005382953.1
	E-value	5.00E-15
	BLAST percent identity	32

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	246
--	------------------------	-----

Protein identifiers	PCC 6912	UYC_07752
	PCC 9212	Chlorogloopsis_PCC9212_joined_4189
	PCC 73103	FisPCC73103_1422
	PCC 7414	FisPCC7414_2473
	PCC 7521	Fischerella_sp._PCC7521_4655
Protein length		95
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03703
---------------------	----------	-----------

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 9212	Chlorogloopsis_PCC9212_joined_5459
	PCC 73103	FisPCC73103_1946
	PCC 7414	FisPCC7414_3072
	PCC 7521	Fischerella_sp._PCC7521_2145
Protein length		251
Putative gene product		Dystrophin-1
Protein name		-
COG-number		COG3034
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07112928.1
	E-value	7.00E-38
	BLAST percent identity	79
	BLAST alignment length	80
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08493876.1
	E-value	7.00E-34
	BLAST percent identity	73
	BLAST alignment length	80

Protein identifiers	PCC 6912	UYC_04895
	PCC 9212	Chlorogloopsis_PCC9212_joined_536
	PCC 73103	FisPCC73103_1477
	PCC 7414	FisPCC7414_2165
	PCC 7521	Fischerella_sp._PCC7521_3320

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		726
Putative gene product		Uncharacterized protein Yegl
Protein name		Yegl
COG-number		COG4248
Best hit outside cyanobacteria	Phylum	Acidobacteria
	Organism	Granulicella tundricola
	Accession	ref YP_004210263.1
	E-value	8.00E-75
	BLAST percent identity	44
	BLAST alignment length	335
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01623720.1
	E-value	6.00E-128
	BLAST percent identity	60
	BLAST alignment length	358
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495348.1
	E-value	1.00E-127
	BLAST percent identity	59
	BLAST alignment length	358

Protein identifiers	PCC 6912	UYC_01407
	PCC 9212	Chlorogloopsis_PCC9212_joined_5094
	PCC 73103	FisPCC73103_4231
	PCC 7414	FisPCC7414_893
	PCC 7521	Fischerella_sp._PCC7521_2284
Protein length		536
Putative gene product		probable peptidase
Protein name		-

Dagan, Roettger, *et al.* – Supplementary Material

COG-number		COG0739
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Selenomonas ruminantium subsp. lactilytica TAM6421
	Accession	ref YP_005432377.1
	E-value	4.00E-36
	BLAST percent identity	56
	BLAST alignment length	126
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07113563.1
	E-value	1.00E-76
	BLAST percent identity	43
	BLAST alignment length	374
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08428623.1
	E-value	4.00E-68
	BLAST percent identity	42
	BLAST alignment length	352

Protein identifiers	PCC 6912	UYC_03451
	PCC 9212	Chlorogloopsis_PCC9212_joined_5701
	PCC 73103	FisPCC73103_4186
	PCC 7414	FisPCC7414_1999
	PCC 7521	Fischerella_sp._PCC7521_2558
Protein length		199
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. PCC 7822
	Accession	ref YP_003890139.1
	E-value	4.00E-32
	BLAST percent identity	61
	BLAST alignment length	89
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05038432.1
	E-value	2.00E-28
	BLAST percent identity	50
	BLAST alignment length	103

Protein identifiers	PCC 6912	UYC_02432
	PCC 9212	Chlorogloopsis_PCC9212_joined_1705
	PCC 73103	FisPCC73103_5161
	PCC 7414	FisPCC7414_4600
	PCC 7521	Fischerella_sp._PCC7521_3519
Protein length		747
Putative gene product		putative peptidase
Protein name		LysM
COG-number		COG0739
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Megamonas funiformis YIT 11815
	Accession	ref ZP_09733812.1
	E-value	9.00E-37
	BLAST percent identity	59
	BLAST alignment length	129

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05026841.1
	E-value	2.00E-73
	BLAST percent identity	41
	BLAST alignment length	468
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08426674.1
	E-value	2.00E-63
	BLAST percent identity	36
	BLAST alignment length	500

Protein identifiers	PCC 6912	UYC_02141
	PCC 9212	Chlorogloopsis_PCC9212_joined_1428
	PCC 73103	FisPCC73103_4051
	PCC 7414	FisPCC7414_6892
	PCC 7521	Fischerella_sp._PCC7521_1939
Protein length		209
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05027162.1
	E-value	1.00E-30

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	48
	BLAST alignment length	139
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427539.1
	E-value	5.00E-25
	BLAST percent identity	41
	BLAST alignment length	151

Protein identifiers	PCC 6912	UYC_00019
	PCC 9212	Chlorogloopsis_PCC9212_joined_4282
	PCC 73103	FisPCC73103_5859
	PCC 7414	FisPCC7414_1329
	PCC 7521	Fischerella_sp._PCC7521_978
Protein length		114
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05028678.1
	E-value	2.00E-14
	BLAST percent identity	37
	BLAST alignment length	113
Best hit in non-heterocyst forming cyanobacteria that are not in our	Organism	Microcoleus vaginatus FGP-2

Dagan, Roettger, *et al.* – Supplementary Material

dataset	Accession	ref ZP_08493731.1
	E-value	9.00E-14
	BLAST percent identity	41
	BLAST alignment length	112

Protein identifiers	PCC 6912	UYC_07811
	PCC 9212	Chlorogloopsis_PCC9212_joined_4244
	PCC 73103	FisPCC73103_4861
	PCC 7414	FisPCC7414_4858
	PCC 7521	Fischerella_sp._PCC7521_4855
Protein length		505
Putative gene product		polyketide synthase
Protein name		HetM
COG-number		COG3320
Best hit outside cyanobacteria	Phylum	Gamma-proteobacteria
	Organism	Beggiatoa sp. PS
	Accession	ref ZP_02000235.1
	E-value	3.00E-101
	BLAST percent identity	47
	BLAST alignment length	403
Best hit in non-heterocyst forming cyanobacteria	Organism	Crocospaera watsonii WH 8501
	Accession	ref ZP_00516263.1
	E-value	6.00E-113
	BLAST percent identity	50
	BLAST alignment length	412
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08494736.1
	E-value	5.00E-113
	BLAST percent identity	48
	BLAST alignment length	392

Protein identifiers	PCC 6912	UYC_02621
	PCC 9212	Chlorogloopsis_PCC9212_joined_1889
	PCC 73103	FisPCC73103_3646
	PCC 7414	FisPCC7414_6826
	PCC 7521	Fischerella_sp._PCC7521_2770
Protein length		66
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Trichodesmium erythraeum IMS101
	Accession	ref YP_723513.1
	E-value	1.00E-08
	BLAST percent identity	37
	BLAST alignment length	63
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03076
	PCC 9212	Chlorogloopsis_PCC9212_joined_5929

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 73103	FisPCC73103_7646
	PCC 7414	FisPCC7414_1788
	PCC 7521	Fischerella_sp._PCC7521_1978
Protein length		92
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05038883.1
	E-value	2.00E-09
	BLAST percent identity	50
	BLAST alignment length	54

Protein identifiers	PCC 6912	UYC_04416
	PCC 9212	Chlorogloopsis_PCC9212_joined_4967
	PCC 73103	FisPCC73103_3964
	PCC 7414	FisPCC7414_2049
	PCC 7521	Fischerella_sp._PCC7521_1013

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		810
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3468
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07111729.1
	E-value	2.00E-49
	BLAST percent identity	34
	BLAST alignment length	448
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495589.1
	E-value	4.00E-42
	BLAST percent identity	34
	BLAST alignment length	346

Protein identifiers	PCC 6912	UYC_05164
	PCC 9212	Chlorogloopsis_PCC9212_joined_791
	PCC 73103	FisPCC73103_4065
	PCC 7414	FisPCC7414_2769
	PCC 7521	Fischerella_sp._PCC7521_3696
Protein length		116
Putative gene product		Thylakoid membrane protein Ssl2009
Protein name		-
COG-number		-

Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002375921.1
	E-value	1.00E-10
	BLAST percent identity	53
	BLAST alignment length	104
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08490772.1
	E-value	2.00E-12
	BLAST percent identity	44
	BLAST alignment length	115

Protein identifiers	PCC 6912	UYC_07056
	PCC 9212	Chlorogloopsis_PCC9212_joined_3535
	PCC 73103	FisPCC73103_6473
	PCC 7414	FisPCC7414_6137
	PCC 7521	Fischerella_sp._PCC7521_4752
Protein length		215
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Trichodesmium erythraeum IMS101
	Accession	ref YP_723861.1
	E-value	6.00E-04
	BLAST percent identity	65
	BLAST alignment length	31
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_04859
	PCC 9212	Chlorogloopsis_PCC9212_joined_502
	PCC 73103	FisPCC73103_7561
	PCC 7414	FisPCC7414_3759
	PCC 7521	Fischerella_sp._PCC7521_1503
Protein length		251
Putative gene product		unknown protein
Protein name		RPII
COG-number		COG3409
Best hit outside cyanobacteria	Phylum	Gamma-proteobacteria
	Organism	Stenotrophomonas maltophilia K279a
	Accession	ref YP_001970192.1
	E-value	1.00E-04
	BLAST percent identity	31
	BLAST alignment length	118

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_00134
	PCC 9212	Chlorogloopsis_PCC9212_joined_4389
	PCC 73103	FisPCC73103_4150
	PCC 7414	FisPCC7414_574
	PCC 7521	Fischerella_sp._PCC7521_2313
Protein length		375
Putative gene product		beta-hexosaminidase
Protein name		NagZ
COG-number		COG1472
Best hit outside cyanobacteria	Phylum	Alpha-proteobacteria
	Organism	Beijerinckia indica subsp. indica ATCC 9039
	Accession	ref YP_001832004.1
	E-value	5.00E-95
	BLAST percent identity	44
	BLAST alignment length	350
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05033
	PCC 9212	Chlorogloopsis_PCC9212_joined_668
	PCC 73103	FisPCC73103_1189
	PCC 7414	FisPCC7414_5889
	PCC 7521	Fischerella_sp._PCC7521_3409
Protein length		159
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Synechocystis sp. PCC 6803
	Accession	ref NP_440019.1
	E-value	3.00E-14
	BLAST percent identity	31
	BLAST alignment length	121
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08431062.1

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	2.00E-23
	BLAST percent identity	43
	BLAST alignment length	108

Protein identifiers	PCC 6912	UYC_07428
	PCC 9212	Chlorogloopsis_PCC9212_joined_3883
	PCC 73103	FisPCC73103_2823
	PCC 7414	FisPCC7414_6615
	PCC 7521	Fischerella_sp._PCC7521_2675
Protein length		68
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002378163.1
	E-value	1.00E-10
	BLAST percent identity	53
	BLAST alignment length	47
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08429796.1
	E-value	4.00E-13
	BLAST percent identity	47
	BLAST alignment length	64

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_04909
	PCC 9212	Chlorogloopsis_PCC9212_joined_550
	PCC 73103	FisPCC73103_1460
	PCC 7414	FisPCC7414_2151
	PCC 7521	Fischerella_sp._PCC7521_3311
Protein length		272
Putative gene product		Uncharacterized sugar transferase HI0872 EC=2.-.-.-;
Protein name		ExoY
COG-number		COG2148
Best hit outside cyanobacteria	Phylum	Bacteriodetes
	Organism	Chlorobium limicola DSM 245
	Accession	ref YP_001943853.1
	E-value	3.00E-15
	BLAST percent identity	35
	BLAST alignment length	151
Best hit in non-heterocyst forming cyanobacteria	Organism	Gloeobacter violaceus PCC 7421
	Accession	ref NP_926182.1
	E-value	2.00E-14
	BLAST percent identity	35
	BLAST alignment length	197
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427156.1
	E-value	3.00E-39
	BLAST percent identity	44
	BLAST alignment length	245

Protein identifiers	PCC 6912	UYC_01783
	PCC 9212	Chlorogloopsis_PCC9212_joined_1102
	PCC 73103	FisPCC73103_4777

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7414	FisPCC7414_942
	PCC 7521	Fischerella_sp._PCC7521_1208
Protein length		91
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07112496.1
	E-value	6.00E-11
	BLAST percent identity	61
	BLAST alignment length	46
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08491150.1
	E-value	7.00E-14
	BLAST percent identity	58
	BLAST alignment length	62

Protein identifiers	PCC 6912	UYC_06071
	PCC 9212	Chlorogloopsis_PCC9212_joined_2632
	PCC 73103	FisPCC73103_7480
	PCC 7414	FisPCC7414_3447
	PCC 7521	Fischerella_sp._PCC7521_1695
Protein length		1368

Dagan, Roettger, *et al.* – Supplementary Material

Putative gene product		hypothetical protein
Protein name		-
COG-number		COG1674
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07112661.1
	E-value	5.00E-119
	BLAST percent identity	40
	BLAST alignment length	638
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08492034.1
	E-value	7.00E-121
	BLAST percent identity	36
	BLAST alignment length	691

Protein identifiers	PCC 6912	UYC_03799
	PCC 9212	Chlorogloopsis_PCC9212_joined_7500
	PCC 73103	FisPCC73103_4316
	PCC 7414	FisPCC7414_1500
	PCC 7521	Fischerella_sp._PCC7521_934
Protein length		214
Putative gene product		heterocyst differentiation related protein
Protein name		PatN
COG-number		COG3115

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_00721
	PCC 9212	Chlorogloopsis_PCC9212_joined_6788
	PCC 73103	FisPCC73103_4536
	PCC 7414	FisPCC7414_5757
	PCC 7521	Fischerella_sp._PCC7521_4505
Protein length		117
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07108840.1
	E-value	4.00E-07
	BLAST percent identity	33
	BLAST alignment length	118
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08429099.1
	E-value	1.00E-04
	BLAST percent identity	33
	BLAST alignment length	110

Protein identifiers	PCC 6912	UYC_02993
	PCC 9212	Chlorogloopsis_PCC9212_joined_6011
	PCC 73103	FisPCC73103_3043
	PCC 7414	FisPCC7414_5591
	PCC 7521	Fischerella_sp._PCC7521_4126
Protein length		87
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

cyanobacteria	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset		
	Organism	Moorea producta 3L
	Accession	ref ZP_08426904.1
	E-value	2.00E-29
	BLAST percent identity	61
	BLAST alignment length	77

Protein identifiers	PCC 6912	UYC_06797
	PCC 9212	Chlorogloopsis_PCC9212_joined_3301
	PCC 73103	FisPCC73103_3878
	PCC 7414	FisPCC7414_6544
	PCC 7521	Fischerella_sp._PCC7521_4484
Protein length		
		775
Putative gene product		
		hypothetical protein
Protein name		
		-
COG-number		
		COG3206
Best hit outside cyanobacteria		
	Phylum	Firmicutes
	Organism	Thermoanaerobacter pseudethanolicus ATCC 33223
	Accession	ref YP_001665535.1
	E-value	8.00E-30
	BLAST percent identity	35
	BLAST alignment length	191
Best hit in non-heterocyst forming cyanobacteria		
	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01622251.1
	E-value	5.00E-132
	BLAST percent identity	37
	BLAST alignment length	745

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427157.1
	E-value	1.00E-125
	BLAST percent identity	38
	BLAST alignment length	748

Protein identifiers	PCC 6912	UYC_03660
	PCC 9212	Chlorogloopsis_PCC9212_joined_5500
	PCC 73103	FisPCC73103_1055
	PCC 7414	FisPCC7414_2648
	PCC 7521	Fischerella_sp._PCC7521_4434
Protein length		289
Putative gene product		Protein FraH
Protein name		FraH
COG-number		COG1716
Best hit outside cyanobacteria	Phylum	Chloroflexi
	Organism	Roseiflexus castenholzii DSM 13941
	Accession	ref YP_001433345.1
	E-value	6.00E-08
	BLAST percent identity	40
	BLAST alignment length	80
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07108834.1
	E-value	3.00E-62
	BLAST percent identity	83
	BLAST alignment length	115
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08491872.1
	E-value	1.00E-56

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	77
	BLAST alignment length	115

Protein identifiers	PCC 6912	UYC_01507
	PCC 9212	Chlorogloopsis_PCC9212_joined_5001
	PCC 73103	FisPCC73103_7802
	PCC 7414	FisPCC7414_2895
	PCC 7521	Fischerella_sp._PCC7521_2551
Protein length		172
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3290
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05027947.1
	E-value	2.00E-17
	BLAST percent identity	39
	BLAST alignment length	127
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08431723.1
	E-value	2.00E-10
	BLAST percent identity	35
	BLAST alignment length	43

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_06256
	PCC 9212	Chlorogloopsis_PCC9212_joined_2798
	PCC 73103	FisPCC73103_898
	PCC 7414	FisPCC7414_4717
	PCC 7521	Fischerella_sp._PCC7521_981
Protein length		129
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_06561
	PCC 9212	Chlorogloopsis_PCC9212_joined_3087
	PCC 73103	FisPCC73103_1788
	PCC 7414	FisPCC7414_5006

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_4419
Protein length		1252
Putative gene product		TPR repeat
Protein name		OGT
COG-number		COG0457
Best hit outside cyanobacteria	Phylum	Spirochaetes
	Organism	Brachyspira hyodysenteriae WA1
	Accession	ref YP_002722686.1
	E-value	2.00E-60
	BLAST percent identity	-
	BLAST alignment length	326
Best hit in non-heterocyst forming cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05028484.1
	E-value	0
	BLAST percent identity	38
	BLAST alignment length	978
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495403.1
	E-value	2.00E-166
	BLAST percent identity	36
	BLAST alignment length	938

Protein identifiers	PCC 6912	UYC_07119
	PCC 9212	Chlorogloopsis_PCC9212_joined_3595
	PCC 73103	FisPCC73103_1359
	PCC 7414	FisPCC7414_2095
	PCC 7521	Fischerella_sp._PCC7521_2979
Protein length		268
Putative gene product		Uncharacterized 15.3 kDa protein in frxC 3'region AltName: Full=URF1; AltName: Full=ORF133;

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	35
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01624541.1
	E-value	1.00E-38
	BLAST percent identity	65
	BLAST alignment length	100
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08494396.1
	E-value	4.00E-38
	BLAST percent identity	70
	BLAST alignment length	99

Protein identifiers	PCC 6912	UYC_00216
	PCC 9212	Chlorogloopsis_PCC9212_joined_4467
	PCC 73103	FisPCC73103_2562
	PCC 7414	FisPCC7414_5847
	PCC 7521	Fischerella_sp._PCC7521_5121
Protein length		154
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-

Dagan, Roettger, *et al.* – Supplementary Material

	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08494107.1
	E-value	7.00E-67
	BLAST percent identity	68
	BLAST alignment length	148

Protein identifiers	PCC 6912	UYC_05282
	PCC 9212	Chlorogloopsis_PCC9212_joined_7594
	PCC 73103	FisPCC73103_3565
	PCC 7414	FisPCC7414_1202
	PCC 7521	Fischerella_sp._PCC7521_1907
Protein length		159
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3678
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07111836.1
	E-value	3.00E-21
	BLAST percent identity	44
	BLAST alignment length	135
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08427650.1
	E-value	6.00E-17
	BLAST percent identity	36
	BLAST alignment length	137

Protein identifiers	PCC 6912	UYC_01673
	PCC 9212	Chlorogloopsis_PCC9212_joined_1000
	PCC 73103	FisPCC73103_4538
	PCC 7414	FisPCC7414_5756
	PCC 7521	Fischerella_sp._PCC7521_4504
Protein length		430
Putative gene product		Calpain-B EC=3.4.22.-; AltName: Full=Calcium-activated neutral proteinase B; Short=CANP B; Contains: RecName: Full=Calpain-B catalytic subunit 1; Contains: RecName: Full=Calpain-B catalytic subunit 2;
Protein name		-
COG-number		COG5022
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-heterocyst forming cyanobacteria	Organism	Oscillatoria sp. PCC 6506
	Accession	ref ZP_07109110.1
	E-value	1.00E-21
	BLAST percent identity	66
	BLAST alignment length	64
Best hit in non-heterocyst forming cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08426021.1
	E-value	3.00E-22
	BLAST percent identity	70
	BLAST alignment length	64

Supp Table S3.

Protein families identified as uniquely shared among heterocystous cyanobacteria of Subsection V. Best BLAST hits in other cyanobacteria, and outside cyanobacteria are also cited.

Protein identifiers	PCC 6912	UYC_01519
	PCC 9212	Chlorogloopsis_PCC9212_joined_4988 Chlorogloopsis_PCC9212_joined_4990 Chlorogloopsis_PCC9212_joined_3441
	PCC 73103	FisPCC73103_3052 FisPCC73103_3051
	PCC 7414	FisPCC7414_2925 FisPCC7414_2926
	PCC 7521	Fischerella_sp._PCC7521_4692
Protein length		288
Putative gene product		Biofilm PIA synthesis N-acetylglucosaminyltransferase icaA EC=2.4.-.-; AltName: Full=Intercellular adhesion protein A;
Protein name		YibD
COG-number		COG1215
Best hit outside cyanobacteria	Phylum	Chloroflexi
	Organism	Chloroflexus aggregans DSM 9485
	Accession	ref YP_002462890.1
	E-value	4.00E-67
	BLAST percent identity	42
	BLAST alignment length	296
Best hit in non-section V cyanobacteria	Organism	Microcystis aeruginosa NIES-843
	Accession	ref YP_001659599.1
	E-value	2.00E-19
	BLAST percent identity	30
	BLAST alignment length	242
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Synechococcus sp. WH 7805
	Accession	ref ZP_01124006.1
	E-value	4.00E-19

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	30
	BLAST alignment length	284

Protein identifiers	PCC 6912	UYC_01882
	PCC 9212	Chlorogloopsis_PCC9212_joined_1195
	PCC 73103	FisPCC73103_1391
	PCC 7414	FisPCC7414_4785
	PCC 7521	Fischerella_sp._PCC7521_4210 Fischerella_sp._PCC7521_5057
Protein length		388
Putative gene product		Glycogen synthase EC=2.4.1.21; AltName: Full=Starch [bacterial glycogen] synthase;
Protein name		-
COG-number		COG0438
Best hit outside cyanobacteria	Phylum	Bacterioidetes
	Organism	Dysgonomonas mossii DSM 22836
	Accession	ref ZP_08469572.1
	E-value	1.00E-69
	BLAST percent identity	37
	BLAST alignment length	389
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002375894.1
	E-value	2.00E-58
	BLAST percent identity	35
	BLAST alignment length	310
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05035657.1
	E-value	2.00E-109
	BLAST percent identity	45
	BLAST alignment length	377

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_07467
	PCC 9212	Chlorogloopsis_PCC9212_joined_3920
	PCC 73103	FisPCC73103_4029
	PCC 7414	FisPCC7414_7013
	PCC 7521	Fischerella_sp._PCC7521_3859
Protein length		423
Putative gene product		Exopolysaccharide production protein exoQ
Protein name		-
COG-number		COG3307
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Desulfotomaculum gibsoniae DSM 7213
	Accession	ref ZP_09100893.1
	E-value	4.00E-30
	BLAST percent identity	30
	BLAST alignment length	366
Best hit in non-section V cyanobacteria	Organism	Nostoc sp. PCC 7120
	Accession	ref NP_488468.1
	E-value	6.00E-49
	BLAST percent identity	35
	BLAST alignment length	382
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08425808.1
	E-value	9.00E-76
	BLAST percent identity	40
	BLAST alignment length	422

Protein identifiers	PCC 6912	UYC_07791
	PCC 9212	Chlorogloopsis_PCC9212_joined_4225
	PCC 73103	FisPCC73103_4838

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7414	FisPCC7414_4838
	PCC 7521	Fischerella_sp._PCC7521_4832
Protein length		277
Putative gene product		3-ketoacyl-(acyl-carrier-protein) reductase (EC:1.1.1.100)
Protein name		FabG
COG-number		COG1028
Best hit outside cyanobacteria	Phylum	Actinobacteria
	Organism	Rhodococcus jostii RHA1
	Accession	ref YP_704919.1
	E-value	4.00E-42
	BLAST percent identity	34
	BLAST alignment length	276
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001866733.1
	E-value	6.00E-96
	BLAST percent identity	58
	BLAST alignment length	267
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08493769.1
	E-value	1.00E-33
	BLAST percent identity	36
	BLAST alignment length	190

Protein identifiers	PCC 6912	UYC_07455
	PCC 9212	Chlorogloopsis_PCC9212_joined_3907
	PCC 73103	FisPCC73103_4038
	PCC 7414	FisPCC7414_7028
	PCC 7521	Fischerella_sp._PCC7521_3850
Protein length		134

Dagan, Roettger, *et al.* – Supplementary Material

Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nostoc sp. PCC 7120
	Accession	ref NP_487377.1
	E-value	2.00E-17
	BLAST percent identity	41
	BLAST alignment length	133
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_00746
	PCC 9212	Chlorogloopsis_PCC9212_joined_6812
	PCC 73103	FisPCC73103_4612
	PCC 7414	FisPCC7414_3559
	PCC 7521	Fischerella_sp._PCC7521_5208
Protein length		165
Putative gene product		flavo-hemoglobin
Protein name		Hmp
COG-number		COG1017

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Clostridium bartlettii DSM 16795
	Accession	ref ZP_02210845.1
	E-value	1.00E-33
	BLAST percent identity	45
	BLAST alignment length	144
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. ATCC 51142
	Accession	ref YP_001805208.1
	E-value	7.00E-46
	BLAST percent identity	58
	BLAST alignment length	144
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08973260.1
	E-value	7.00E-46
	BLAST percent identity	58
	BLAST alignment length	144

Protein identifiers	PCC 6912	UYC_06091
	PCC 9212	Chlorogloopsis_PCC9212_joined_2651
	PCC 73103	FisPCC73103_1867
	PCC 7414	FisPCC7414_6461
	PCC 7521	Fischerella_sp._PCC7521_5101
Protein length		383
Putative gene product		serine/threonine kinase
Protein name		-
COG-number		COG5263
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Blautia hydrogenotrophica DSM 10507
	Accession	ref ZP_03782649.1
	E-value	1.00E-18

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	52
	BLAST alignment length	81
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001865506.1
	E-value	2.00E-19
	BLAST percent identity	38
	BLAST alignment length	152
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_07376
	PCC 9212	Chlorogloopsis_PCC9212_joined_3836
	PCC 73103	FisPCC73103_2837
	PCC 7414	FisPCC7414_2728
	PCC 7521	Fischerella_sp._PCC7521_3964
Protein length		655
Putative gene product		serine/threonine protein kinase
Protein name		PknD
COG-number		COG0515
Best hit outside cyanobacteria	Phylum	Acidobacteria
	Organism	Candidatus Chloracidobacterium thermophilum B
	Accession	ref YP_004863306.1
	E-value	7.00E-60
	BLAST percent identity	44
	BLAST alignment length	278
Best hit in non-section V cyanobacteria	Organism	Nostoc sp. PCC 7120

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	ref NP_488406.1
	E-value	9.00E-132
	BLAST percent identity	64
	BLAST alignment length	310
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Arthrospira platensis str. Paraca
	Accession	ref ZP_06384094.1
	E-value	7.00E-108
	BLAST percent identity	51
	BLAST alignment length	351

Protein identifiers	PCC 6912	UYC_05139
	PCC 9212	Chlorogloopsis_PCC9212_joined_765
	PCC 73103	FisPCC73103_4791
	PCC 7414	FisPCC7414_1806
	PCC 7521	Fischerella_sp._PCC7521_517
Protein length		989
Putative gene product		Phosphate-binding protein pstS Short=PBP; Flags: Precursor;
Protein name		-
COG-number		COG0226
Best hit outside cyanobacteria	Phylum	Firmicutes
	Organism	Symbiobacterium thermophilum IAM 14863
	Accession	ref YP_074064.1
	E-value	2.00E-21
	BLAST percent identity	31
	BLAST alignment length	231
Best hit in non-section V cyanobacteria	Organism	Anabaena variabilis ATCC 29413
	Accession	ref YP_321376.1
	E-value	2.00E-120
	BLAST percent identity	70
	BLAST alignment length	286

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Cylindrospermopsis raciborskii CS-505
	Accession	ref ZP_06308877.1
	E-value	7.00E-74
	BLAST percent identity	51
	BLAST alignment length	257

Protein identifiers	PCC 6912	UYC_05175
	PCC 9212	Chlorogloopsis_PCC9212_joined_802
	PCC 73103	FisPCC73103_883
	PCC 7414	FisPCC7414_6213
	PCC 7521	Fischerella_sp._PCC7521_645
Protein length		294
Putative gene product		Elongation factor G Short=EF-G;
Protein name		-
COG-number		COG0523
Best hit outside cyanobacteria	Phylum	Chlamydiae
	Organism	Pedosphaera parvula Ellin514
	Accession	ref ZP_03629449.1
	E-value	3.00E-123
	BLAST percent identity	61
	BLAST alignment length	294
Best hit in non-section V cyanobacteria	Organism	Prochlorococcus marinus str. MIT 9211
	Accession	ref YP_001551350.1
	E-value	1.00E-34
	BLAST percent identity	35
	BLAST alignment length	307
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Synechococcus sp. WH 8016
	Accession	ref ZP_08955941.1
	E-value	3.00E-32

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	35
	BLAST alignment length	320

Protein identifiers	PCC 6912	UYC_00646
	PCC 9212	Chlorogloopsis_PCC9212_joined_6715
	PCC 73103	FisPCC73103_3008
	PCC 7414	FisPCC7414_5039
	PCC 7521	Fischerella_sp._PCC7521_1534
Protein length		50
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Anabaena variabilis ATCC 29413
	Accession	ref YP_325424.1
	E-value	4.00E-15
	BLAST percent identity	62
	BLAST alignment length	50
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_05283
	PCC 9212	Chlorogloopsis_PCC9212_joined_7595
	PCC 73103	FisPCC73103_6009
	PCC 7414	FisPCC7414_4621
	PCC 7521	Fischerella_sp._PCC7521_3078
Protein length		150
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3678
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002380009.1
	E-value	3.00E-12
	BLAST percent identity	48
	BLAST alignment length	69
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01856
	PCC 9212	Chlorogloopsis_PCC9212_joined_1171
	PCC 73103	FisPCC73103_7121
	PCC 7414	FisPCC7414_1162

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_4917
Protein length		190
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Trichodesmium erythraeum IMS101
	Accession	ref YP_721775.1
	E-value	1.00E-04
	BLAST percent identity	27
	BLAST alignment length	215
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03986
	PCC 9212	Chlorogloopsis_PCC9212_joined_7323
	PCC 73103	FisPCC73103_7674
	PCC 7414	FisPCC7414_603
	PCC 7521	Fischerella_sp._PCC7521_4572
Protein length		67
Putative gene product		hypothetical protein

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002380165.1
	E-value	6.00E-09
	BLAST percent identity	50
	BLAST alignment length	62
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05176
	PCC 9212	Chlorogloopsis_PCC9212_joined_803
	PCC 73103	FisPCC73103_884
	PCC 7414	FisPCC7414_6212
	PCC 7521	Fischerella_sp._PCC7521_644
Protein length		143
Putative gene product		Protein archease AltName: Full=Protein ZBTB8OS;
Protein name		-
COG-number		COG1371
Best hit outside cyanobacteria	Phylum	Gamma-proteobacteria

Dagan, Roettger, *et al.* – Supplementary Material

	Organism	Nitrosococcus halophilus Nc4
	Accession	ref YP_003528400.1
	E-value	4.00E-43
	BLAST percent identity	59
	BLAST alignment length	143
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02593
	PCC 9212	Chlorogloopsis_PCC9212_joined_1861
	PCC 73103	FisPCC73103_2632
	PCC 7414	FisPCC7414_7163
	PCC 7521	Fischerella_sp._PCC7521_818
Protein length		128
Putative gene product		RNA pyrophosphohydrolase EC=3.6.1.-; AltName: Full=(Di)nucleoside polyphosphate hydrolase;
Protein name		-
COG-number		COG0494
Best hit outside cyanobacteria	Phylum	Planctomycetes
	Organism	Blastopirellula marina DSM 3645
	Accession	ref ZP_01093308.1
	E-value	2.00E-35

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	50
	BLAST alignment length	129
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01015
	PCC 9212	Chlorogloopsis_PCC9212_joined_4580
	PCC 73103	FisPCC73103_3786
	PCC 7414	FisPCC7414_7722
	PCC 7521	Fischerella_sp._PCC7521_2658
Protein length		85
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01549
	PCC 9212	Chlorogloopsis_PCC9212_joined_878
	PCC 73103	FisPCC73103_7504
	PCC 7414	FisPCC7414_7660
	PCC 7521	Fischerella_sp._PCC7521_3890
Protein length		496
Putative gene product		Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase EC=2.-.-.; AltName: Full=Undecaprenyl phosphate-alpha-L-Ara4N transferase; AltName: Full=4-amino-4-deoxy-L-arabinose lipid A transferase; AltName: Full=Polymyxin resistance protein pmrK;
Protein name		-
COG-number		COG1807
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_06197
	PCC 9212	Chlorogloopsis_PCC9212_joined_2747
	PCC 73103	FisPCC73103_1723
	PCC 7414	FisPCC7414_5145
	PCC 7521	Fischerella_sp._PCC7521_1951
Protein length		252
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG0739
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7424
	Accession	ref YP_002375715.1
	E-value	2.00E-17
	BLAST percent identity	47
	BLAST alignment length	90
Best hit in non-section V cyanobacteria	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

that are not in our dataset	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01343
	PCC 9212	Chlorogloopsis_PCC9212_joined_5156
	PCC 73103	FisPCC73103_810
	PCC 7414	FisPCC7414_1194
	PCC 7521	Fischerella_sp._PCC7521_1021
Protein length		137
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Synechococcus sp. PCC 7002
	Accession	Synechococcus sp. PCC 7002
	E-value	7.00E-13
	BLAST percent identity	34
	BLAST alignment length	137
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Cyanothece sp. ATCC 51472
	Accession	ref ZP_08975139.1
	E-value	9.00E-10
	BLAST percent identity	33
	BLAST alignment length	135

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_01632
	PCC 9212	Chlorogloopsis_PCC9212_joined_959
	PCC 73103	FisPCC73103_6299
	PCC 7414	FisPCC7414_1057
	PCC 7521	Fischerella_sp._PCC7521_2092
Protein length		108
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03497
	PCC 9212	Chlorogloopsis_PCC9212_joined_5655

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 73103	FisPCC73103_1218
	PCC 7414	FisPCC7414_1433
	PCC 7521	Fischerella_sp._PCC7521_1075
Protein length		105
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Microcoleus chthonoplastes PCC 7420
	Accession	ref ZP_05025787.1
	E-value	7.00E-08
	BLAST percent identity	33
	BLAST alignment length	101
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03404
	PCC 9212	Chlorogloopsis_PCC9212_joined_5741
	PCC 73103	FisPCC73103_6194
	PCC 7414	FisPCC7414_4091
	PCC 7521	Fischerella_sp._PCC7521_3451

Dagan, Roettger, *et al.* – Supplementary Material

Protein length		317
Putative gene product		Conserved virulence factor C
Protein name		-
COG-number		COG5635
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7425
	Accession	ref YP_002485728.1
	E-value	1.00E-08
	BLAST percent identity	28
	BLAST alignment length	220
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_00614
	PCC 9212	Chlorogloopsis_PCC9212_joined_6686
	PCC 73103	FisPCC73103_6572
	PCC 7414	FisPCC7414_5404
	PCC 7521	Fischerella_sp._PCC7521_412
Protein length		95
Putative gene product		hypothetical protein
Protein name		-
COG-number		-

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02799
	PCC 9212	Chlorogloopsis_PCC9212_joined_2059
	PCC 73103	FisPCC73103_5819
	PCC 7414	FisPCC7414_2932
	PCC 7521	Fischerella_sp._PCC7521_3598
Protein length		53
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nodularia spumigena CCY9414
	Accession	ref ZP_01629899.1
	E-value	4.00E-16
	BLAST percent identity	78
	BLAST alignment length	45
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_06504
	PCC 9212	Chlorogloopsis_PCC9212_joined_3037
	PCC 73103	FisPCC73103_4090
	PCC 7414	FisPCC7414_7682
	PCC 7521	Fischerella_sp._PCC7521_2822
Protein length		427
Putative gene product		O-acetyltransferase OatA EC=2.3.1.-;
Protein name		-
COG-number		COG1835
Best hit outside cyanobacteria	Phylum	Epsilon-proteobacteria
	Organism	Campylobacter gracilis RM3268
	Accession	ref ZP_05625095.1
	E-value	1.00E-27
	BLAST percent identity	43
	BLAST alignment length	167

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_01551
	PCC 9212	Chlorogloopsis_PCC9212_joined_879
	PCC 73103	FisPCC73103_7505
	PCC 7414	FisPCC7414_7659
	PCC 7521	Fischerella_sp._PCC7521_3891
Protein length		680
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG1807
Best hit outside cyanobacteria	Phylum	Chloroflexi
	Organism	Caldilinea aerophila DSM 14535 = NBRC 104270
	Accession	ref YP_005443184.1
	E-value	9.00E-49
	BLAST percent identity	30
	BLAST alignment length	615
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05395
	PCC 9212	Chlorogloopsis_PCC9212_joined_7701
	PCC 73103	FisPCC73103_5378
	PCC 7414	FisPCC7414_4665
	PCC 7521	Fischerella_sp._PCC7521_1665
Protein length		45
Putative gene product		Photosystem I reaction center subunit IX
Protein name		PsaJ
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Anabaena variabilis ATCC 29413
	Accession	ref YP_321995.1
	E-value	5.00E-16
	BLAST percent identity	82
	BLAST alignment length	44
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Raphidiopsis brookii D9
	Accession	ref ZP_06304113.1

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	6.00E-16
	BLAST percent identity	80
	BLAST alignment length	45

Protein identifiers	PCC 6912	UYC_03608
	PCC 9212	Chlorogloopsis_PCC9212_joined_5552
	PCC 73103	FisPCC73103_2168
	PCC 7414	FisPCC7414_3133
	PCC 7521	Fischerella_sp._PCC7521_4626
Protein length		319
Putative gene product		Leukotoxin secretion protein D
Protein name		LktD
COG-number		COG0845
Best hit outside cyanobacteria	Phylum	Alpha-proteobacteria
	Organism	Rhodobacteraceae bacterium KLH11
	Accession	ref ZP_05125090.1
	E-value	4.00E-11
	BLAST percent identity	30
	BLAST alignment length	162
Best hit in non-section V cyanobacteria	Organism	Acaryochloris marina MBIC11017
	Accession	ref YP_001518311.1
	E-value	5.00E-19
	BLAST percent identity	36
	BLAST alignment length	171
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495096.1
	E-value	4.00E-27
	BLAST percent identity	40
	BLAST alignment length	170

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_01208
	PCC 9212	Chlorogloopsis_PCC9212_joined_5280
	PCC 73103	FisPCC73103_4019
	PCC 7414	FisPCC7414_7001
	PCC 7521	Fischerella_sp._PCC7521_3868
Protein length		283
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03958
	PCC 9212	Chlorogloopsis_PCC9212_joined_7350
	PCC 73103	FisPCC73103_5108

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7414	FisPCC7414_2562
	PCC 7521	Fischerella_sp._PCC7521_5270
Protein length		407
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02968
	PCC 9212	Chlorogloopsis_PCC9212_joined_6032
	PCC 73103	FisPCC73103_7534
	PCC 7414	FisPCC7414_3640
	PCC 7521	Fischerella_sp._PCC7521_1676
Protein length		556

Dagan, Roettger, *et al.* – Supplementary Material

Putative gene product		two-component hybrid sensor and regulator
Protein name		PHYA
COG-number		COG0642
Best hit outside cyanobacteria	Phylum	Delta-proteobacteria
	Organism	Corallocooccus coralloides DSM 2259
	Accession	ref YP_005366365.1
	E-value	4.00E-63
	BLAST percent identity	30
	BLAST alignment length	501
Best hit in non-section V cyanobacteria	Organism	Nostoc sp. PCC 7120
	Accession	ref NP_484590.1
	E-value	3.00E-104
	BLAST percent identity	68
	BLAST alignment length	247
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Synechococcus sp. PCC 7335
	Accession	ref ZP_05036830.1
	E-value	6.00E-80
	BLAST percent identity	47
	BLAST alignment length	293

Protein identifiers	PCC 6912	UYC_02891
	PCC 9212	Chlorogloopsis_PCC9212_joined_1975
	PCC 73103	FisPCC73103_1296
	PCC 7414	FisPCC7414_2901
	PCC 7521	Fischerella_sp._PCC7521_1099
Protein length		123
Putative gene product		hypothetical protein
Protein name		-
COG-number		-

Dagan, Roettger, *et al.* – Supplementary Material

Best hit outside cyanobacteria	Phylum	Gamma-proteobacteria
	Organism	Nitrococcus mobilis Nb-231
	Accession	ref ZP_01125770.1
	E-value	2.00E-26
	BLAST percent identity	48
	BLAST alignment length	122
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_06801
	PCC 9212	Chlorogloopsis_PCC9212_joined_3305
	PCC 73103	FisPCC73103_6526
	PCC 7414	FisPCC7414_3062
	PCC 7521	Fischerella_sp._PCC7521_3727
Protein length		269
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	Nitrospirae
	Organism	Candidatus Nitrospira defluvii
	Accession	ref YP_003796251.1
	E-value	1.00E-13

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	30
	BLAST alignment length	200
Best hit in non-section V cyanobacteria	Organism	Cyanothece sp. PCC 7425
	Accession	ref YP_002481823.1
	E-value	2.00E-30
	BLAST percent identity	37
	BLAST alignment length	195
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Acaryochloris sp. CCME 5410
	Accession	ref ZP_09251532.1
	E-value	7.00E-30
	BLAST percent identity	33
	BLAST alignment length	201

Protein identifiers	PCC 6912	UYC_03653
	PCC 9212	Chlorogloopsis_PCC9212_joined_5507
	PCC 73103	FisPCC73103_1048
	PCC 7414	FisPCC7414_6014
	PCC 7521	Fischerella_sp._PCC7521_544
Protein length		549
Putative gene product		UPF0159 protein CT_632
Protein name		-
COG-number		COG1351
Best hit outside cyanobacteria	Phylum	Chloroflexi
	Organism	Herpetosiphon aurantiacus DSM 785
	Accession	ref YP_001547062.1
	E-value	0.00E+00
	BLAST percent identity	65
	BLAST alignment length	549
Best hit in non-section V cyanobacteria	Organism	-

Dagan, Roettger, *et al.* – Supplementary Material

	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_00915
	PCC 9212	Chlorogloopsis_PCC9212_joined_4676
	PCC 73103	FisPCC73103_3564
	PCC 7414	FisPCC7414_1200
	PCC 7521	Fischerella_sp._PCC7521_1905
Protein length		295
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	Bacterioidetes
	Organism	Spirosoma linguale DSM 74
	Accession	ref YP_003387456.1
	E-value	3.00E-28
	BLAST percent identity	29
	BLAST alignment length	272
Best hit in non-section V cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01620944.1
	E-value	2.00E-69
	BLAST percent identity	46
	BLAST alignment length	248

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Arthrospira maxima CS-328
	Accession	ref ZP_03273171.1
	E-value	6.00E-49
	BLAST percent identity	39
	BLAST alignment length	250

Protein identifiers	PCC 6912	UYC_04947
	PCC 9212	Chlorogloopsis_PCC9212_joined_589
	PCC 73103	FisPCC73103_1483
	PCC 7414	FisPCC7414_2174
	PCC 7521	Fischerella_sp._PCC7521_3326
Protein length		137
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001867542.1
	E-value	2.00E-09
	BLAST percent identity	66
	BLAST alignment length	38
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_03455
	PCC 9212	Chlorogloopsis_PCC9212_joined_5697
	PCC 73103	FisPCC73103_4497
	PCC 7414	FisPCC7414_1991
	PCC 7521	Fischerella_sp._PCC7521_2562
Protein length		183
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001866410.1
	E-value	1.00E-25
	BLAST percent identity	49
	BLAST alignment length	98
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08491991.1
	E-value	1.00E-20
	BLAST percent identity	47
	BLAST alignment length	96

Dagan, Roettger, *et al.* – Supplementary Material

Protein identifiers	PCC 6912	UYC_01216
	PCC 9212	Chlorogloopsis_PCC9212_joined_5275
	PCC 73103	FisPCC73103_4876
	PCC 7414	FisPCC7414_2231
	PCC 7521	Fischerella_sp._PCC7521_1595
Protein length		567
Putative gene product		serine/threonine kinase
Protein name		PknB
COG-number		COG0515
Best hit outside cyanobacteria	Phylum	Acidobacteria
	Organism	Candidatus Chloracidobacterium thermophilum B
	Accession	ref YP_004863306.1
	E-value	4.00E-42
	BLAST percent identity	40
	BLAST alignment length	263
Best hit in non-section V cyanobacteria	Organism	Nodularia spumigena CCY9414
	Accession	ref ZP_01628558.1
	E-value	3.00E-119
	BLAST percent identity	65
	BLAST alignment length	269
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Raphidiopsis brookii D9
	Accession	ref ZP_06305755.1
	E-value	1.00E-91
	BLAST percent identity	55
	BLAST alignment length	267

Protein identifiers	PCC 6912	UYC_05291
	PCC 9212	Chlorogloopsis_PCC9212_joined_7603
	PCC 73103	FisPCC73103_6633
	PCC 7414	FisPCC7414_3035

Dagan, Roettger, *et al.* – Supplementary Material

	PCC 7521	Fischerella_sp._PCC7521_767
Protein length		279
Putative gene product		hypothetical protein
Protein name		-
COG-number		COG3468
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	'Nostoc azollae' 0708
	Accession	ref YP_003720818.1
	E-value	8.00E-42
	BLAST percent identity	54
	BLAST alignment length	122
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Moorea producta 3L
	Accession	ref ZP_08428905.1
	E-value	2.00E-17
	BLAST percent identity	38
	BLAST alignment length	127

Protein identifiers	PCC 6912	UYC_01840
	PCC 9212	Chlorogloopsis_PCC9212_joined_1155
	PCC 73103	FisPCC73103_7707
	PCC 7414	FisPCC7414_4146
	PCC 7521	Fischerella_sp._PCC7521_2992
Protein length		315
Putative gene product		hypothetical protein

Dagan, Roettger, *et al.* – Supplementary Material

Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	Beta-proteobacteria
	Organism	Dechloromonas aromatica RCB
	Accession	ref YP_285573.1
	E-value	4.00E-76
	BLAST percent identity	42
	BLAST alignment length	307
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_05597
	PCC 9212	Chlorogloopsis_PCC9212_joined_7017
	PCC 73103	FisPCC73103_3684
	PCC 7414	FisPCC7414_5244
	PCC 7521	Fischerella_sp._PCC7521_4377
Protein length		182
Putative gene product		Uncharacterized protein sll0944
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-

Dagan, Roettger, *et al.* – Supplementary Material

	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Lyngbya sp. PCC 8106
	Accession	ref ZP_01622390.1
	E-value	4.00E-19
	BLAST percent identity	56
	BLAST alignment length	71
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	Microcoleus vaginatus FGP-2
	Accession	ref ZP_08495243.1
	E-value	9.00E-19
	BLAST percent identity	48
	BLAST alignment length	81

Protein identifiers	PCC 6912	UYC_05169
	PCC 9212	Chlorogloopsis_PCC9212_joined_796
	PCC 73103	FisPCC73103_1507
	PCC 7414	FisPCC7414_4263
	PCC 7521	Fischerella_sp._PCC7521_2463
Protein length		323
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

Dagan, Roettger, *et al.* – Supplementary Material

	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001864797.1
	E-value	4.00E-33
	BLAST percent identity	44
	BLAST alignment length	249
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_02554
	PCC 9212	Chlorogloopsis_PCC9212_joined_1821
	PCC 73103	FisPCC73103_4061
	PCC 7414	FisPCC7414_3435
	PCC 7521	Fischerella_sp._PCC7521_1926
Protein length		65
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nodularia spumigena CCY9414
	Accession	ref ZP_01629346.1

Dagan, Roettger, *et al.* – Supplementary Material

	E-value	2.00E-14
	BLAST percent identity	59
	BLAST alignment length	59
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_07587
	PCC 9212	Chlorogloopsis_PCC9212_joined_4035
	PCC 73103	FisPCC73103_3368
	PCC 7414	FisPCC7414_3520
	PCC 7521	Fischerella_sp._PCC7521_4133
Protein length		79
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	Nostoc punctiforme PCC 73102
	Accession	ref YP_001864091.1
	E-value	3.00E-10
	BLAST percent identity	63
	BLAST alignment length	75

Dagan, Roettger, *et al.* – Supplementary Material

Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-

Protein identifiers	PCC 6912	UYC_06859
	PCC 9212	Chlorogloopsis_PCC9212_joined_3357
	PCC 73103	FisPCC73103_5497
	PCC 7414	FisPCC7414_7451
	PCC 7521	Fischerella_sp._PCC7521_3275
Protein length		84
Putative gene product		hypothetical protein
Protein name		-
COG-number		-
Best hit outside cyanobacteria	Phylum	-
	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-
	BLAST alignment length	-
Best hit in non-section V cyanobacteria that are not in our dataset	Organism	-
	Accession	-
	E-value	-
	BLAST percent identity	-

	BLAST alignment length	-
--	------------------------	---